10/634,895

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/40, 15/86, C07K 14/18, A61K 39/12, C12N 7/01, 7/00, 5/10 — A1

(11) International Publication Number: WO 96/40933

(43) International Publication Date: 19 December 1996 (19.12.96)

(21) International Application Number: PCT/US96/09209

(22) International Filing Date: 6 June 1996 (06.06.96)

(30) Priority Data:
08/483,292 7 June 1995 (07.06.95) US

(71) Applicants: THE GOVERNMENT OF THE UNITED STATES OF AMERICA, represented by THE SECRETARY DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; National Institute of Health, Office of TechnologyTransfer, Suite 305, 6011 Executive Boulevard, Rockville, MD 20852 (US). MAHIDOL UNIVERSITY AT SALAYA [TH/TH]; 25/25 Phutthamonthon 4, Nakhonpathom 73170 (TH).

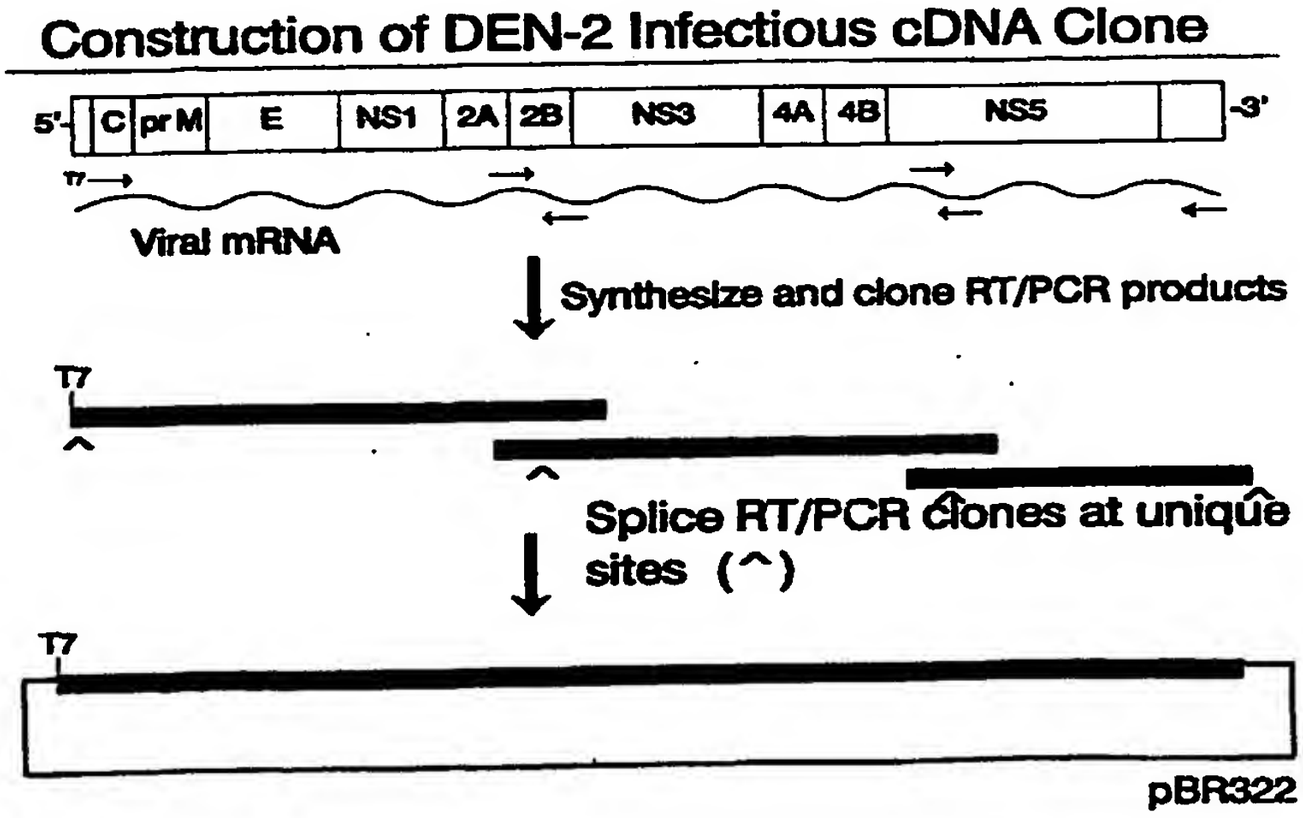
(72) Inventors: BHAMARAPRAVATI, Natth; Mahidol University at Salaya, Nakhonchaisri, Nakhonpathom 73170 (TH). BUTRAPET, Siritom; 396/27 Soi Watsuna Naram, Bangkok-Noi, Bangkok 10700 (TH). CHANG, Jeffrey; 4237 Beaver Creek, Fort Collins, CO 80526 (US). GUBLER, Duane, J.; 717 Dartmouth Trail, Fort Collins, CO 80525 (US). HALSTEAD, Scott, B.; 5901 Montrose Road, Rockville, MN 20852 (US). KINNEY, Richard; 2914 Brumbaugh Drive, Fort Collins, CO 80526 (US). TRENT, Dennis, W.; 8002 Lions Crest Way, Gaithersburg, MD 20879 (US).

(74) Agents: SPRATT, Gwendolyn, D. et al.; Needle & Rosenberg, Suite 1200, 127 Peachtree Street, N.E., Atlanta, GA 30303-1811 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: INFECTIOUS DENGUE 2 VIRUS PDK-53 AS QUADRAVALENT VACCINE



(57) Abstract

The invention relates to infectious cDNA clones for Dengue 2 virus, strain 16681, and its live, attenuated vaccine derivative, PDK-53 (DEN-2 PDK-53). The invention also relates to infectious cDNA clones for chimeric viruses characterized as expressing structural genes of a Dengue 1, Dengue 3, or Dengue 4 attenuated virus in the context of the nonstructural genes of the Dengue 2 PDK-53 virus (DEN-2/1, DEN-2/3, DEN-2/4). The invention further relates to genetic constructs encoding these cDNAs, and host cells containing these constructs. The invention moreover relates to quadravalent vaccines providing immunity against all four serotypes of dengue virus comprising DEN-2 PDK-53 infectious clone derivative, DEN-2/1, DEN-2/3, or DEN-2/4 viruses, and related methods of immunization.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

INFECTIOUS DENGUE 2 VIRUS PDK-53 AS QUADRAVALENT VACCINE

Field of the Invention

The invention relates to infectious cDNA clones for Dengue 2 virus, strain 16681, and its live, attenuated vaccine derivative, PDK-53 (DEN-2 PDK-53). The invention also relates to infectious cDNA clones for chimeric viruses characterized as expressing structural genes of a Dengue 1, Dengue 3, or Dengue 4 attenuated virus in the context of the nonstructural genes of the Dengue 2 PDK-53 virus (DEN-2/1, DEN-2/3, DEN-2/4). The invention further relates to genetic constructs encoding these cDNAs, and host cells containing these constructs. The invention moreover relates to quadravalent vaccines providing immunity against all four serotypes of dengue virus comprising DEN-2 PDK-53 infectious clone derivative, DEN-2/1, DEN-2/3, or DEN-2/4 viruses, and related methods of immunization.

Background of the Invention

Arthropod-borne viruses (arboviruses) are a diverse group of viruses that have been lumped together on the basis of their ecological niche, which involves cycles of transmission between vertebrate hosts and arthropod vectors such as mosquitos and ticks. The prototype arbovirus is yellow fever virus, a flavivirus, which was isolated in 1927. In the 1950s, the Rockefeller Foundation established a number of field stations in

various tropical countries for the purpose of isolating new viruses. The 1985 International Catalogue of Arboviruses Including Certain Other Viruses of Vertebrates contains registrations for 504 discrete arboviruses, 124 of which have caused disease in humans. Thirty-four viruses of the Flavivirus genus (family Flaviviridae) of arboviruses are human pathogens (Karabatsos, 1985). (All publications cited hereunder are incorporated herein by reference.)

According to a 1992 World Health Organization (WHO) press release (Press Release WHO/74, November 24, 1992), dengue hemorrhagic fever is one of the most important and increasing mosquito-transmitted infections in the world, with more than 85 countries in Asia, the Pacific Islands, Africa, Central America, and South America being threatened with dengue outbreaks. Dengue fever was known in the past as "breakbone fever" due to the severe muscular and joint pain that accompanied the high fever during this infection. Dengue is an under-reported disease: it is thought that millions of cases occur each year.

Dengue (DEN) viruses, which are flaviviruses, are classified antigenically into 4 serotypes (DEN-1, DEN-2, DEN-3, and DEN-4). Multiple serotypes are now endemic in most countries in the tropics. DEN viruses are transmitted to humans principally by *Aedes aegypti* mosquitos throughout much of the tropical and subtropical region of the world. Viruses of all four serotypes infect humans and cause clinically inapparent infection or illness ranging from dengue fever to severe and often

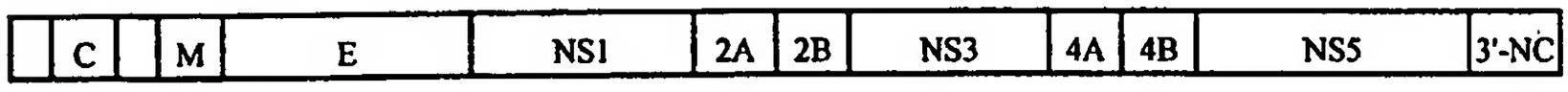
fatal dengue hemorrhagic fever/dengue shock syndrome (DHF/DSS). DHF/DSS has been associated epidemiologically and experimentally with immune enhancement of virus replication by preexisting, subneutralizing levels of heterotypic antibody. About 90% or more of patients with DHF/DSS are children who are 14 years old or younger (Halstead, 1970; Halstead, 1988). Case fatality rates in untreated individuals can be as high as 15-20%. Between 1956 and 1978, hospitalization of more than 350,000 dengue patients and about 12,000 deaths in Southeast Asia were reported to the WHO (Halstead, 1980). More recent dengue epidemics in Asia, the Pacific islands, the Americas, and Africa indicate that the incidence, with up to 40 million cases annually, and geographic distribution of the disease is increasing in *Aedes aegypti*-infested areas of the world (Halstead, 1984; Gubler, 1988; Brandt, 1990).

Since eradication of *Aedes aegypti* mosquitos appears to be practically infeasible, development of safe, effective vaccines against all four serotypes of DEN virus is a WHO priority (Gubler, 1988; Brandt, 1988; Brandt, 1990). Since the level of DEN virus replication in certified cell cultures yields insufficient antigenic mass to produce effective inactivated vaccines, priorities are given to developing effective live, attenuated vaccine viruses and using a variety of expression systems such as recombinant vaccinia or avipox virus (live vaccine), recombinant baculovirus (subunit vaccine), and recombinant *E. coli* (subunit vaccine) to express certain genes of the DEN viral genome (Brandt, 1988; Brandt, 1990).

Flaviviruses are enveloped RNA viruses 45 to 50 nm in diameter that contain a single-stranded, positive-sense capped RNA genome of approximately 11 kb. The RNA genome does not have a 3'-terminal poly(A) tail. Because the genetic molecule of flaviviruses is positive or messenger RNA (mRNA)-sense, naked genomic RNA injected, transfected, or electroporated into mammalian or invertebrate cells is capable of associating directly with the ribosomal protein synthetic machinery of the cell. All of the viral proteins are translated from the inserted viral genomic mRNA. These virus-specified proteins then replicate the viral genome, resulting in intracellular virus maturation and release of infectious virus from the transfected cell.

The gene organization of the flavivirus mRNA genome, illustrated below, is 5'-noncoding region (5'-NC)-capsid-premembrane/membrane (prM/M)-envelope (E)-nonstructural protein 1 (NS1)-NS2A-NS2B-NS3-NS4A-NS4B-NS5-3'-noncoding region (3'-NC). The structural proteins capsid, prM/M, and E and nonstructural proteins are translated as a large precursor polyprotein molecule from a single long open reading frame in the mRNA genome. The individual mature viral proteins are processed from the polyprotein by both cell and virus specified proteases (Westaway *et al.*, 1985; Coia *et al.*, 1988; Speight and Westaway, 1989; Rice *et al.*, 1985).

Genome Organization of Dengue Virus and Other Flaviviruses



The structural proteins are those viral proteins that are incorporated into the mature virion. The virion consists of an icosahedral capsid (C) that packages the viral genomic mRNA (nucleocapsid). The nucleocapsid is surrounded by a cell-derived lipid membrane into which the envelope (E) and mature membrane (M) proteins are imbedded. The virus-specific nonstructural genes, NS1-NS5, are expressed in the cytoplasm of the infected cell and are involved in the replication and maturation of the viral RNA genome and viral proteins.

The E glycoprotein of the virus is exposed to the environment and is involved in attachment and entry of the virus into the cell. The E protein is the primary viral immunogen against which the infected vertebrate host develops virus-specific neutralizing antibody. The E gene is the most common target for development of molecular systems to express the encoded E glycoprotein. However, immunization with various purified nonstructural genes of the virus have been shown to elicit protective immunity against challenge with wild-type virus, probably via cytotoxic T-cell mediated lysis of infected cells which express viral nonstructural proteins on the cell surface.

Vaccination can be one of the most cost effective ways to prevent dengue fever and DHF/DSS. Since 1979 the WHO has supported research on dengue vaccine development at the Mahidol University in Bangkok, Thailand (Press Release WHO/74, November 24, 1992). Investigators at Mahidol University have developed four live, attenuated candidate vaccine viruses, one for each of the four serotypes, by serial passage of the virulent parent

viruses in primary dog kidney (PDK) or fetal rhesus lung (FRhL) cell culture (Yoksan *et al.*, 1986; Bhamarapravati *et al.*, 1987). Phase 1 and Phase 2 clinical trials in Thailand have demonstrated that the vaccine is both safe and immunogenic in humans. The vaccines now need to be tested for efficacy in large numbers of children (Press Release WHO/74, November 24, 1992). To preclude the possible severe DHF/DSS immune enhancement phenomenon in vaccinees who might be infected naturally with a heterologous serotype of wild-type DEN virus following immunization with a single serotype of vaccine virus, it is essential that humans be vaccinated with a quadravalent vaccine to provide immunity against all four serotypes of the virus.

Summary of the Invention

The invention provides a quadravalent vaccine providing immunity against all four serotypes of dengue virus comprising a DEN-2 PDK-53 infectious clone-derived virus.

The invention also provides a quadravalent vaccine providing immunity against all four serotypes of dengue virus comprising a chimeric DEN-2/1 virus.

The invention further provides a quadravalent vaccine providing immunity against all four serotypes of dengue virus comprising a chimeric DEN-2/3 virus.

The invention moreover provides a quadravalent vaccine providing immunity against all four serotypes of dengue virus comprising a chimeric DEN-2/4 virus.

The invention additionally provides a quadravalent vaccine providing immunity against all four serotypes of dengue virus comprising DEN-2 PDK-53 infectious clone-derived and chimeric DEN-2/1, DEN-2/3, and DEN-2/4 viruses.

In another aspect, the invention provides a method of immunization in which a desired immune response is produced against all four serotypes of dengue virus comprising the step of administering to a subject a quadravalent vaccine comprising DEN-2 PDK-53 infectious clone-derived and chimeric DEN-2/1, DEN-2/3, and DEN-2/4 viruses.

In yet another aspect, the invention provides a composition of matter comprising a full genome-length infectious cDNA clone for a DEN-2 virus, strain 16681.

The invention also provides a composition of matter comprising a full genome-length infectious cDNA clone for a DEN-2 virus of a strain characterized as replicating to high titer in cell culture.

The invention further provides a composition of matter comprising a full genome-length infectious cDNA clone for a DEN-2 virus, strain 16681, having the identifying characteristics of ATCC 69826.

In still another aspect, the invention provides a composition of matter comprising a full genome-length infectious cDNA clone for a DEN-2 virus, strain 16681, attenuated derivative, PDK-53.

The invention also provides a composition of matter comprising a full genome-length infectious cDNA clone for a DEN-2 virus attenuated derivative,

characterized as replicating to high titer in cell culture.

The invention further provides a composition of matter comprising a full genome-length infectious cDNA clone for a DEN-2 virus, strain 16681, attenuated derivative, PDK-53, having the identifying characteristics of ATCC 69825.

In another aspect, the invention provides a composition of matter comprising a full genome-length infectious cDNA clone of a chimeric DEN-2/1 virus, wherein the virus is characterized as expressing the prM and E genes of a DEN-1 attenuated virus in the context of the nonstructural genes of the DEN-2 PDK-53 virus. The DEN-1 attenuated virus may be DEN-1 PDK-13.

The invention also provides a composition of matter comprising a full genome-length infectious cDNA clone of a chimeric DEN-2 virus, wherein the virus is characterized as expressing the antigenicity of a DEN-1 attenuated virus.

In yet another aspect, the invention provides a composition of matter comprising a full genome-length infectious cDNA clone of a chimeric DEN-2/3 virus, wherein the virus is characterized as expressing the prM and E genes of a DEN-3 attenuated virus in the context of the nonstructural genes of the DEN-2 PDK-53 virus. The DEN-3 attenuated virus may be DEN-3 PGMK30/FRhL-3.

The invention also provides a composition of matter comprising a full genome-length infectious cDNA clone of a chimeric DEN-2 virus, wherein the virus is

characterized as expressing the antigenicity of a DEN-3 attenuated virus.

In still another aspect, the invention provides a composition of matter comprising a full genome-length infectious cDNA clone of a chimeric DEN-2/4 virus, wherein the virus is characterized as expressing the prM and E genes of a DEN-4 attenuated virus in the context of the nonstructural genes of the DEN-2 PDK-53 virus. The DEN-4 attenuated virus may be DEN-4 PDK-48.

The invention also provides a composition of matter comprising a full genome-length infectious cDNA clone of a chimeric DEN-2 virus, wherein the virus is characterized as expressing the antigenicity of a DEN-4 attenuated virus.

Additionally, the invention provides a genetic construct comprising a DNA sequence operably encoding the polyprotein of DEN-2 virus, strain 16681. The polyprotein may be the polyprotein encoded by the nucleotide sequence of SEQ ID NO:1.

The invention also provides a genetic construct comprising a DNA sequence operably encoding at least one protein of DEN-2 virus, strain 16681. The protein may be a protein encoded by the nucleotide sequence of SEQ ID NO: 1.

Further, the invention provides a genetic construct comprising a DNA sequence operably encoding the polyprotein of DEN-2 virus, strain 16681, attenuated derivative, PDK-53. The polyprotein may be the polyprotein encoded by the nucleotide sequence of SEQ ID NO:2.

The invention also provides a genetic construct comprising a DNA sequence operably encoding at least one protein of DEN-2 virus, strain 16681, attenuated derivative, PDK-53. The protein may be a protein encoded by the nucleotide sequence of SEQ ID NO: 2.

Moreover, the invention provides a genetic construct comprising a DNA sequence operably encoding at least one structural protein of DEN-1 PDK-13. The structural protein may be a structural protein encoded by the nucleotide sequence of SEQ ID NO: 124.

In another aspect, the invention provides a genetic construct comprising a DNA sequence operably encoding at least one structural protein of DEN-3 PGMK30/FRhL-3. The structural protein may be a structural protein encoded by the nucleotide sequence of SEQ ID NO: 125.

In still another aspect, the invention provides a genetic construct comprising a DNA sequence operably encoding at least one structural protein of DEN-4 PDK-48. The structural protein may be a structural protein encoded by the nucleotide sequence of SEQ ID NO: 126.

In yet another aspect, the invention includes a host cell comprising any of the above genetic constructs.

Brief Description of the Drawings

Figure 1: Strategy for construction of the full genome-length cDNA clone of DEN-2 virus. Using PCR technology, cDNA is amplified from the genomic RNA of the virus and cloned. Subclones are spliced together at unique, overlapping restriction enzyme sites to construct

the full genome-length clone. Numbered arrows upstream (right arrows) and downstream (left primers used to amplify the cDNA in PCR reactions.

Figure 2: Transcription of genomic mRNA from the full-length infectious cDNA clone of DEN-2 virus. The recombinant plasmid is linearized at the unique XbaI site at the 3'-end of the genomic cDNA. Bacteriophage T7 RNA polymerase recognizes the T7 promoter engineered at the 5'-end of the cDNA and transcribes full-length viral mRNA from the cDNA template.

Figure 3: Restriction enzyme sites identified in the nucleotide sequence of the RNA genome of DEN-2 16681 virus. Locations for the sites are indicated by the genome nucleotide numbers. Restriction enzymes that cleave the DEN-2 genomic cDNA at only a single location are listed vertically at the top of the figure. The resolution of the RENZ graph is 97.5 nucleotides per dot.

Figure 4: Growth curve of DEN-2 16681 virus in C6/36 mosquito cells.

Figure 5: **(A)** Polaroid prints showing RT/PCR amplification of the entire mRNA genome of DEN-2 virus, strain 16681, in the form of 5 cDNA amplicons. The molecular weight marker (MW) consists of linear, double-stranded DNA markers of various base pair (bp) lengths. The top 2 gels show 5-μl aliquots of the original RT/PCR reactions. The bottom two gels show 10% of the yield following HMC agarose gel purification of the remaining 95-μl reaction aliquots. **(B)** Primers (amplimers) used in the RT/PCR reactions and the expected sizes of the resulting cDNA amplicons.

Figure 6: EcoRI restriction enzyme digests of F2, F2-Sal, Sal-F2, and F3 miniprep recombinant plasmid DNA. Plasmids from individual colonies resulting from transformation with independent ligated, recombinant plasmid molecules are numbered. The insert in the single F2-8 plasmid was too small and was discarded. The remaining recombinant plasmids contained cDNA inserts of expected size. As expected, F2-Sal cDNA contained two internal EcoRI sites; the Sal-F2 and F3 plasmids contained a single internal EcoRI site. EcoRI digestion of the recombinant plasmids regenerated linearized, wild-type 3.9-kb pCRII vector. For an undetermined reason, one of the EcoRI sites in plasmid F3-1 did not cut.

Figure 7: Schematic diagram showing the genomic locations of DEN-2 16681 virus-specific cDNA clones. Clones indicated with asterisks were spliced together at the indicated restriction enzyme sites to construct the full genome-length cDNA clone. Black horizontal bars indicate clone regions that were sequenced. Light gray regions of horizontal bars indicate clone regions that were not sequenced.

Figure 8: (A) Effect of adding Taq extender reagent to PCR reactions. The 5.2-kbp amplicon of St. Louis encephalitis virus was readily obtained by extended PCR (+) but not by standard PCR (-). (B) Agarose gel electropherogram showing DEN-2 PDK-53 F1, F2, and F3 amplicons derived by extended PCR.

Figure 9: Schematic diagram showing the genome locations of errors identified in the cDNA clones of DEN-2 16681. Errors are indicated by short vertical tick marks.

Figure 10: Schematic diagram illustrating the approximate genome locations of the nucleotide discrepancies between the data of Applicants and those of Blok *et al.* (1992) for the sequence of the genome of DEN-2 virus, strain 16681.

Figure 11: Nucleotide sequence of the genome of DEN-2 strain 16681 virus. Differences between the data determined by Blok *et al.* (1992) (DEN-2-16681.BLOK) and those obtained by Applicants (DEN-2-16681.RK). The genome nucleotide positions of the sequence differences are listed vertically. The solid squares indicate those nucleotide differences that also encode amino acid substitutions. The remaining nucleotide differences are either silent, encoding the same amino acid, or lie within the 5'-noncoding (5'-NC) or 3'-noncoding region (3'-NC).

Figure 12: Schematic diagram showing the DEN-2 PDK-53 virus-specific cDNA clones and the approximate locations of cDNA errors (vertical tick marks) identified by nucleotide sequence analyses. Clones marked with an asterisk were used in the construction of the DEN-2 PDK-53 virus-specific full-length cDNA clone. Clone #19 had a 203-bp deletion (horizontal line).

Figure 13: Schematic summary of the DEN-2 16681 vs. PDK-53 virus sequencing projects. Arrows indicate the nucleotide differences detected between the two genomes. Triangles indicate those nucleotide changes that resulted in amino acid substitutions.

Figure 14: Finalized nucleotide and amino acid sequence of the RNA genome of DEN-2 virus, strain 16681 (SEQ ID NO:1). The nucleotide and amino acid mutations

that were determined to have occurred in DEN-2 virus, strain PDK-53, are indicated at the appropriate positions (SEQ ID NO:2). The EcoRI, SstI, MulI,and T7 promoter sites that were engineered immediately preceding the 5'-terminal nucleotide of the virus-specific genomic cDNA are shown. The start positions of the viral genes and noncoding regions (5'-NC and 3'-NC) are shown. Potential sites of Asn-linked glycosylation (Asn-X-Ser or Thr, where X = any amino acid) in prM, E, and NS1 are indicated by asterisks. The deduced amino acid sequence is indicated in standard single-letter abbreviation: A = Ala, C = Cys, D = Asp, E = Glu, F = Phe, G = Gly, H = His, I = Ile, K = Lys, L = Leu, M = Met, N = Asn, P = Pro, Q = Gln, R = Arg, S = Ser, T = Thr, V = Val, W = Trp, Y = Tyr.

Figure 15: Construction of intermediate clone F2 by ligating the F2-Sal SphI/HpaI fragment and Sal-F2 HpaI/KpnI fragment into pUC18. The resulting F2 clone contained a nonsilent cDNA error at genome nucleotide position 1730.

Figure 16: Correction of the intermediate F2 clone. A new PCR amplicon was cloned and sequenced. The SphI/HpaI fragment of this clone was spliced into F2 to construct F2-C having the correct nucleotide at genome position 1730.

Figure 17: Construction of the intermediate F1/3/4/5 cDNA clone for DEN-2 16681 virus. The thick solid black bars indicate DEN-2 virus-specific cDNA, illustrated with the RENZ sites of the MCS of the plasmid. The RENZ sites used in each step of the splicing strategy are indicated in underlined, bold characters. The top half of the

figure shows construction of F1/3/4/5-pUC18. The bottom portion of the figure illustrates the making of F1/3/4/5-pUC19. The final step in the construction of the full genome-length cDNA clone involved the ligation of the F2-C SphI/KpnI cDNA fragment into plasmid containing cDNA F1/3/4/5 and cut with RENZs SphI/KpnI. Although F2-C cDNA could not be cloned into F1/3/4/5-pUC18, it was readily cloned into F1/3/4/5-pUC19. The pUC18 plasmid containing a small insert of cDNA made for Venezuelan equine encephalitis (VEE) virus was used simply to move F1 and F4/5 into pUC18 in a 3-molecule ligation reaction. The VEE virus-specific cDNA was spliced out during this process. Arrowheads under cDNA bars indicate orientation of mRNA-sense cDNA strand.

Figure 18: Orientation specific cloning of full genome-length cDNA of DEN-2 16681 virus into the multiple cloning site of pUC19. Although the full-length cDNA was readily cloned in pUC19, multiple attempts to insert the cDNA into pUC18 failed. Presumably, interaction of the cDNA with pUC18-specific gene transcripts, translation of a toxic DEN-2 polypeptide, or translation of a toxic pUC18/DEN-2 fusion polypeptide produced deleterious effects in *E. coli*. Large arrows indicate orientation of mRNA-sense cDNA strands in the pUC plasmid backbone. Smaller arrows indicate orientations of the *lac Z* and ampicillin genes as well as the origin of replication. DEN-2 insert is indicated by a thick solid black line.

Figure 19: Insertion of the MCS of plasmid pUC19 into pBR322 in both orientations to construct pBRUC-138 and pBRUC-139. The pUC18 HindIII (blunt-ended = BL)/EcoRI

MCS fragment was ligated into pBR322 cut with AvaI (BL)/EcoRI to construct pBRUC-138. The pUC18 EcoRI (BL)/HindIII MCS fragment was ligated into pBR322 cut with AvaI (BL)/HindIII to make pBRUC-139. In both cases, the tetracycline gene of pBR322 was removed. pBRUC-138 = 2992-bp (61-bp MCS + 2931-bp pBR322 deletion vector). pBRUC-139 = 3022-bp (61-bp MCS + 2961-bp pBR322 deletion vector). Orientations of ORI, ROP, and the Amp gene are indicated.

Figure 20: Construction of pD2/IC-30P, the full genome-length cDNA clone of DEN-2 16681 virus, in plasmid pBR322 (pBRUC-139 (SphI-) derivative). The F3/4/5 clone cDNA was ligated into pBRUC-139 first (Top of Figure), followed by F1-E and F2-C. Viable, infectious DEN-2 virus was successfully obtained from viral mRNA transcribed from this clone.

Figure 21: Construction of pD2/IC-130V, the full genome-length cDNA clone of DEN-2 PDK-53 virus. A nonsilent error in cDNA clone F3-3C was corrected by splicing in a correct BstBI/NheI fragment from clone F3.5-6 (Top). The resulting corrected clone F3-3CC was spliced into the 16681 F345-F clone in pBRUC-139. cDNA fragments F1-79B, F2-16B, and the recombinant F3/4/5 vector DNA were spliced together in a single ligation reaction to produce pD2/IC-130V. The NheI site occurs at genome nucleotide position 6646. Therefore, the PDK-53 virus-specific full-length cDNA clone contains the parental 16681 virus-specific nucleotide at position 8571. This nucleotide difference is silent; it does not encode an amino acid change. Other than the 8571 position, DEN-2 16681 and

PDK-53 viruses are identical in nucleotide sequence from nucleotide position 6646 to the 3' terminus of the genome.

Figure 22: Agarose gel electropherogram of viral genomic mRNA extracted from gradient-purified, wild-type DEN-2 16681 virus and Venezuelan equine encephalitis (VEE) virus. The quantity of RNA loaded onto the gel ranged from 22 ng to 383 ng. The stock RNA was quantitated spectrophotometrically at 260 nm. The genome-length RNA band is clearly visible between the 4153-bp and 6788-bp MW marker bands. Bands were visualized by incorporating 200 ng/ml of ethidium bromide stain in the gel and electrophoresis buffer.

Figure 23: Transcription of RNA from pVE/IC-92 (VEE virus clone) and pD2/IC-20 (DEN-2 16681 virus clone). Transcription reaction conditions (100 ng linearized DNA template, 12.5 mM DTT, 2.7 u/μl RNasin, 0.15 mM NTPs, 3.3 U/μl T7 RNA polymerase (Stratagene) in commercial buffer (Stratagene)) yielded high quantity and quality of infectious mRNA transcripts from the pVE/IC-92 clone and 3'-end truncation products of that clone. However, these reaction conditions failed to permit transcription of RNA from the pD2/IC-20 clone or two of its 3'-end transcription products (clone linearized at the NsiI or MroI site instead of at the 3'-terminal XbaI site). pVE/IC-92 plasmid linearized at the MluI (3'-terminal), SphI, Tth111I, HindIII, SalI, and StuI sites in the cDNA clone yielded RNA transcripts of 11447, 11377, 7541, 2407, 1620, and 674 base length, respectively (the more intense, prominent bands in these gel lanes).

Figure 24: Transcription of RNA from the DEN-2 16681 cDNA clone pD2/IC-20. **(A)** Transcription of RNA using different quantities of linearized plasmid template (a,b). The cap analog m7G(5')ppp(5')A was not included in the reaction. **(B)** Transcription of 5'-capped RNA with inclusion of cap analog in the reaction. Transcription was accomplished with the Ampliscribe transcription kit from Epicentre Technologies. T7 pol = bacteriophage T7 RNA polymerase.

Figure 25: Transcription of full genome-length, infectious viral mRNA from XbaI-linearized DEN-2 16681 plasmid pD2/IC-30P (A and D replicate clones resulting from independent bacterial colonies transformed with the recombinant pBRUC/DEN-2 plasmid) and PDK-53 plasmid pD2/IC-130V (F and J replicates). Genomic "viral RNA" extracted from gradient-purified wild-type DEN-2 16681 virus was electrophoresed in lanes 2 and 10. Aliquots of transcription reactions sampled before (T7 RNA polymerase "-") and after (T7 Pol "+") addition of T7 RNA polymerase are shown. Only the linearized plasmid DNA template is observed in the absence of the polymerase.

Figure 26: Transcription of RNA from pD2/IC-20, pD2/IC-30P, and pD2/IC-130V in the presence or absence of T7 RNA polymerase or cap analog in the transcription reaction. All lanes shown are on a single gel. Transcription was performed with the Ampliscribe transcription kit.

Figure 27: Derivation tree for the construction of the DEN-2 16681 and PDK-53 virus-specific full genome-length cDNA clones pD2/IC-30P and pD2/IC-130V,

respectively, and chimeric 16681/PDK-53 clones derived from the two prototype clones.

Figure 28: Genotype maps of DEN-2 16681 and PDK-53 virus-specific full genome-length cDNAs and their chimeric derivatives. The scale at the top indicates relative genome nucleotide position in thousands. The graph resolution is 119.1444 bp/dot. cDNA regions contributed by the parental DEN-2 16681 virus are indicated by solid black bars. Regions derived from the DEN-2 PDK-53 vaccine virus are indicated by stippled bars. The 8 mutations identified by sequence analyses of the genomes of the 16681 and PDK-53 viruses are indicated. The virus-specific 5-noncoding nucleotides are indicated in lower case characters. The amino acids encoded by the virus-specific nucleotide mutations in the protein coding region of the genome are indicated in upper case, single-letter amino acid abbreviation.

Figure 29: Results of spot-sequencing PCR amplicons amplified from seed stocks of viruses derived from full genome-length cDNA clones. Dots indicate nucleotide sequence identity to the DEN-2 16681 virus. The expected virus-specific nucleotides for the genotype of each virus are shown. Those nucleotide positions that have actually been confirmed by sequence analysis are indicated by underlined nucleotide base characters. The actual genome nucleotide positions are indicated at the bottom of the Figure.

Figure 30: Recombinant full-length pD2/IC-30P-A and pD2/IC-130V-F plasmids extracted from 1-ml aliquots of *E. coli* TB-1 cultures submitted to ATCC.

Figure 31: Partial nucleotide sequences of candidate vaccine viruses:

DEN-1 16007 PDK-13 (D1.VAC) (SEQ ID NO: 124)

DEN-2 16681 PDK-53 (D2.VAC) (see SEQ ID NO: 2)

DEN-3 16562 PGMK-30/FRhL-3 (D3.VAC) (SEQ ID NO: 125)

DEN-4 1036 PDK-48 (D4.VAC) (SEQ ID NO: 126)

aligned with the nucleotide and deduced amino acid sequences of DEN-2 16681 virus (see SEQ ID NO:1). Dots in the DEN-1, DEN-3, and DEN-4 sequences signify identity with the DEN-2 sequence.

Figure 32: Partial amino acid sequences of candidate vaccine viruses:

DEN-1 16007 PDK-13 (D1.VAC) (SEQ ID NO: 124)

DEN-2 16681 PDK-53 (D2.VAC) (see SEQ ID NO: 2)

DEN-3 16562 PGMK-30/FRhL-3 (D3.VAC) (SEQ ID NO: 125)

DEN-4 1036 PDK-48 (D4.VAC) (SEQ ID NO: 126)

aligned with the deduced amino acid sequence of DEN-2 16681 virus (see SEQ ID NO:1). Dots in the DEN-1, DEN-3, and DEN-4 sequences signify identity with the DEN-2 sequence.

Figure 33: Mutagenesis analysis of the 5' end of the prM gene. The 447-452 sequence ("AACCAC" in DEN-2) can be mutated to "CTCGAG" in all four DEN viruses to create a XhoI site for cassette splicing. This modification results in conservative Thr-Thr to Ser-Ser substitutions at amino acid positions prM 4-5 in DEN-2 virus. By creating this XhoI site, all four viruses will contain the sequence FHLSSR at amino acid positions prM 1-6 (see Figure 32). Nucleotide mutations that are necessary to create the XhoI site are indicated by bold, underlined

characters in the nucleotide sequences of D2.VAC, D1.VAC, D3.VAC, and D4.VAC and their respective primers designed for amplification in PCR.

Figure 34: Mutagenesis analysis of the 3' end of the E gene. The 2344-2349 sequence ("TCACGC" in DEN-2) can be mutated to "TCTAGA" in all four DEN viruses to create a XbaI site for cassette splicing. This modification results in no amino acid change in DEN-2 at this site, but substitutions do occur in the other three viruses. By creating this XhoI site, all four viruses will contain the sequence SRS at amino acid positions E 470-472 (see Figure 32). Nucleotide mutations that are necessary to create the XbaI site are indicated by bold, underlined characters in the nucleotide sequences of D2.VAC, D1.VAC, D3.VAC, and D4.VAC and their respective primers designed for amplification in PCR.

Figure 35: Construction of DEN-2 PDK-53 cassette plasmids pF1-Xho and pF2-Xba. **(A)** pF1-Xho: Clone PCR cDNA amplicons F1-prM5' and F1-prM3' into TA-vector. Sequence and splice correct clones together at the SphI site in the TA-vector to construct pF1-prM53 (not shown). Subclone the prM53 cDNA into SstI/SphI-cut pF1-E (see Figure 20) to construct pF1-Xho. **(B)** pF2-Xba: Clone PCR cDNA amplicons F2-E5' and F2-E3' into TA-vector. Splice correct clones together at the XbaI site in the TA-vector to construct pF2-E53 (not shown). Subclone the SphI/HpaI E53 cDNA fragment into pF2-16B (see Figure 21), which itself is subcloned into pBRUC-139 between the SphI/KpnI sites (not shown), to construct pF2-Xho. PCR amplimer designations are underlined. Solid black bars indicate newly

synthesized and sequence-characterized cDNA. Stippled bar indicates previously synthesized cDNA. Graph resolution = 64.1857 nucleotides/dot.

Figure 36: Construction of chimeric plasmids containing the prM and E genes (XhoI-XbaI cDNA fragment) of DEN-1, DEN-3, or DEN-4 candidate vaccine virus within the genetic background of DEN-2 PDK-53 virus. pD2V-CAS12 was constructed by ligating the SstI/SphI fragment of pF1-Xho and SphI/KpnI fragment of pF2-Xba (see Figure 33) into a truncated form of pD2/IC-130V (see Figure 21). pD2/IC-130V was truncated by restricting the full-length clone at the NsiI-4696 and 3'-end XbaI sites, blunt-ending with T4 DNA polymerase, and religating. This procedure removed genome nucleotides 4696-10723, thereby removing the XhoI-5426 and 3'-end XbaI sites, which would otherwise interfere with construction of chimeric plasmid cassettes using XhoI and XbaI sites. The cassette strategy employs PCR amplification of DEN-1, DEN-3, and DEN-4 cDNAs containing the prM and E genes; cutting the amplicons with XhoI/XbaI; cloning resulting fragments into pD2V-CAS12 to construct pD1V-CAS12, pD3V-CAS12, and pD4V-CAS12 chimeric cassettes; confirming the chimeric XhoI/XbaI insert by nucleotide sequence analysis; and then subcloning the SstI/KpnI fragment of the chimeric cassette into pD2/IC-130V to construct the chimeric full genome-length cDNA clones from which chimeric DEN-2/1, -2/3, and -2/4 viruses are derived. The genetic background of DEN-2 PDK-53 virus is illustrated by the solid black bars. The heterologous DEN-1, DEN-3, and DEN-4 cDNA inserts are indicated by the stippled bars. The pBRUC-139 plasmid backbone is not

illustrated for pD1V-CAS12, pD3V-CAS12, or pD4V-CAS12 chimeric plasmid. Resolution = 110.5464 bp/dot.

Detailed Description of the Invention

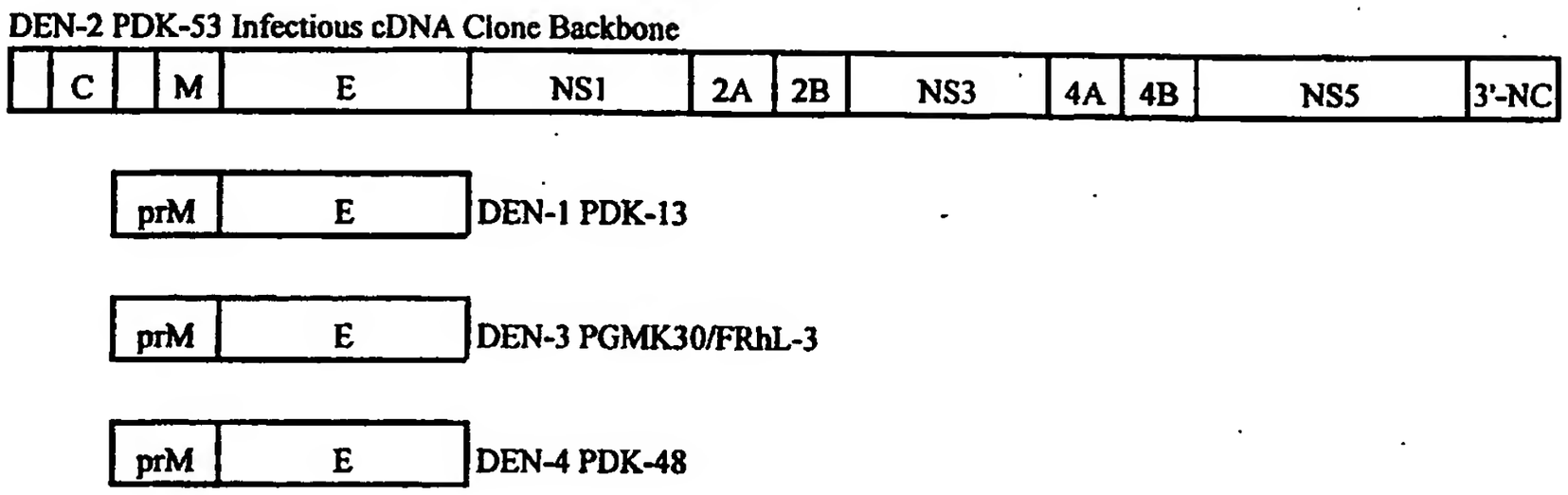
We developed a quadravalent vaccine by initially constructing a full genome-length infectious cDNA clone for DEN-2 virus. We chose serotype 2 of DEN virus because virus strains of this serotype generally replicate to high titer in cell culture. We chose to develop an infectious clone for the 16681 strain of DEN-2 virus because the candidate vaccine viruses developed by Mahidol University are currently the best live, attenuated vaccine virus candidates in terms of immunogenic efficacy and lack of reactogenicity in vaccinees. We developed an infectious cDNA clone of the 16681 strain, which is the parent to the DEN-2 PDK-53 candidate vaccine virus developed at Mahidol University, to permit engineering of second and later generation live, attenuated DEN vaccine viruses.

The infectious clone strategy was initiated with the virulent parental 16681 strain obtained from the Division of Vector-Borne Infectious Diseases (DVBID) of the Centers for Disease Control and Prevention (CDC) virus collection. We synthesized cDNA from the DEN-2 16681 viral RNA. The immediate objective was to obtain an accurate full genome-length infectious cDNA clone of the 16681 strain of DEN-2 virus, since it was essential to develop a reliable experimental system to permit routine genetic engineering of the cDNA and recovery of virus. Our approach involved using polymerase chain reaction (PCR) technology to create cDNA clones that could be spliced together to construct a

single full genome-length clone (Figure 1) from which full-length, infectious DEN-2 genomic mRNA could be transcribed (Figure 2).

The first full-length sequence-characterized cDNA clone, designated pD2/IC-20, was constructed in the high copy number pUC19 plasmid vector. Successful transcription of genome-length DEN-2 16681 viral RNA from pD2/IC-20 was clearly demonstrated by agarose gel electrophoresis of the transcription reaction product. However, RNA transcribed from this particular clone failed to yield infectious virus. It was determined that cDNA errors had occurred during the clone manipulations. We then decided to reconstruct the full-length clone in the low copy number pBR322 plasmid. The full-length cDNA of DEN-2 16681 virus was successfully moved into pBR322 to construct pD2/IC-30P. Full-length, infectious DEN-2 16681 genomic RNA was subsequently transcribed from pD2/IC-30P.

The DEN-1 PDK-13, DEN-2 PDK-53, DEN-3 PGMK-30/FRhL-3, and DEN-4 PDK-48 vaccine viruses were obtained from Mahidol University. Our goal involved replacement of the entire genomic cDNA backbone of the DEN-2 16681 full-length clone with the cognate cDNA cloned from the genome of the DEN-2 PDK-53 candidate vaccine virus. The prM and E genes of the DEN-2 PDK-53 virus are then replaced with the prM and E genes of the DEN-1 PDK-13, DEN-3 PGMK30/FRhL-3, and DEN-4 PDK-48 candidate vaccine viruses to construct chimeric DEN-2/1, DEN-2/3, and DEN-2/4 viruses containing the nonstructural genes of the DEN-2 PDK-53 virus and the prM and E genes of the heterologous DEN viruses.



It is contemplated that chimeric, infectious clone-derived DEN-2/1, DEN-2/3, and DEN-2/4 viruses will result in immediate improvement in the efficacy of a quadravalent vaccine. Our preliminary data from Mahidol University indicate that very small amounts of the DEN-2 PDK-53 vaccine virus were required to infect and immunize humans. However, the DEN-1, DEN-3, and DEN-4 vaccine virus candidates had approximately 30-fold to 2000-fold lower infectivity for humans. The low infective efficacies of the DEN-1, DEN-3, and DEN-4 viruses create significant problems in terms of vaccine efficacy in eliciting seroconversion in vaccinees, as well as problems of vaccine production for mass vaccination programs, since a large volume, up to 1 ml, of undiluted cell culture-derived vaccine virus must be administered to achieve even minimal levels of infectivity for these viruses. Since the increased infectivity of the DEN-2 PDK-53 vaccine virus is likely due to more efficient virus replication, and since this replicative efficacy is controlled by the nonstructural proteins of the virus, then chimeric vaccine viruses that express the relevant immunogenic structural proteins of DEN-1, DEN-3, or DEN-4 virus in the context of replication control by the nonstructural gene products of

the DEN-2 PDK-53 virus should replicate better and be more infective and immunogenic in human vaccinees than the original DEN-1, DEN-3, and DEN-4 vaccine viruses containing nonchimeric genotypes.

A quadravalent vaccine is obtained upon completion of the following steps:

(1) A full genome-length infectious cDNA clone for a DEN-2 virus, strain 16681, is constructed.

(2) A full genome-length infectious cDNA clone for a DEN2-16681 attenuated derivative, PDK-53, is constructed, preferably by substituting the genomic cDNA backbone of the DEN2-16681 full length clone with the corresponding cDNA cloned from the genome of the DEN-2 PDK-53 candidate vaccine virus.

(3) The candidate DEN-1, DEN-3, and DEN-4 vaccine viruses are subjected to PCR amplification of cDNA from extracted genomic RNA, and chimeric infectious cDNA clones expressing the prM and E genes of DEN-1, DEN-3, and DEN-4 viruses, respectively, in the context of the nonstructural genes of the DEN-2 PDK-53 virus are constructed.

(4) The infectious clone-derived chimeric DEN-2/1, DEN-2/3, and DEN-2/4 vaccine viruses are tested to ensure that they:

(a) Are viable;

(b) Express appropriate virus-specific immunogens;

(c) Replicate to sufficient titer in cell culture;

(d) Are infectious and immunogenic for humans; and

(e) Retain phenotypic markers of attenuation.

There is no good animal model for investigating dengue pathogenesis. DEN viruses are naturally transmitted between mosquitos and humans. Although lower primates can be infected with these viruses, they do not develop the clinical profiles that occur in humans. Infectious clone-derived viruses can be compared to their more virulent parental strains using certain *in vitro* and *in vivo* markers:

In Vitro Markers:

Plaque size in cell culture;

Temperature sensitivity;

Cytopathic effects (CPE) in LLC-MK_2 cells; and

Replication in macrophages.

In Vivo Markers:

Virulence by intracranial route in mice;

Viremia in monkeys;

Virulence by intracranial route in monkeys; and

Elicitation of neutralizing antibodies in animals.

Infectious cDNA clones are expressed, the resulting RNA transcripts are transfected into permissible cells, and the live, attenuated viruses are formulated into vaccines.

Additionally, the DEN-2 PDK-53 and chimeric DEN-2/1, DEN-2/3, and DEN-2/4 infectious cDNA clones can by themselves confer immunity by DNA immunization, a form of gene therapy involving the direct inoculation of naked DNA into the host such that its expression produces an immune response (*e.g.*, Ulmer *et al.*, 1993 (DNA immunization protected against influenza); Cox *et al.*, 1993 (DNA immunization protected against herpesvirus); Xiang *et al.*, 1994 (DNA immunization protected against rabies); Sedegah *et al.*, 1994 (DNA immunization protected against malaria)).

Moreover, infectious cDNA clones are exquisite tools for studying the molecular biology of virus structure, function, and replication. This has been amply demonstrated for many RNA viruses in the literature, including Venezuelan equine encephalitis virus as reported by Kinney *et al.* (1989). A successful infectious cDNA clone of DEN-2 virus permits important investigations of

dengue virus replication, pathogenesis, and antigenic structure. Infectious clone cDNA templates permit the *directed* engineering of virus vaccines. Directed site-specific, nonrandom mutations can readily be made in infectious cDNA clones, and therefore in clone-derived viruses, using a wide variety of DNA modification enzymes, restriction endonucleases, and *in vitro* mutagenesis methods. DNA is easier to manipulate than RNA, and the 10^{-9} error rate of DNA replication is much lower than the 10^{-3} - 10^{-4} error rate produced by RNA polymerases. Infectious cDNA clones permit direct analyses of the phenotypic effects of individual and cumulative mutations in the viral genome. An infectious cDNA clone provides a "gold standard" reference sequence for a vaccine.

Particular aspects of the invention may be more readily understood by reference to the following examples, which are intended to exemplify the invention, without limiting its scope to the particular exemplified embodiments.

EXAMPLES

Information:

Most of the background, protocols, and recipes used in recombinant DNA work can be found in *Molecular Cloning: A Laboratory Manual* (Sambrook *et al.*, 1989), and *Current Protocols in Molecular Biology* (Ausubel *et al.*, 1989).

Viruses:

The virulent parental DEN-2 16681 strain was immediately available in the DVBID collection of viruses. We received the DEN-1 PDK-13, DEN-2 PDK-53, DEN-3 PGMK-30/FRhL-3, and DEN-4 PDK-48 vaccine viruses from Mahidol University. The DEN vaccine viruses were passaged in primary dog kidney (PDK) cells because this cell culture is included among those cell types that are certified for human use by the Bureau of Biologics, US Food and Drug Administration (Yoksan *et al.*, 1986). The virus strain designations are shown below:

Virus	Parent Strain	Vaccine Derivative Strain
DEN-1	16007	PDK-13
DEN-2	16681	PDK-53
DEN-3	16562	PGMK-30/FRhL-3

DEN-4 1036 PDK-48

PDK = primary dog kidney cells

FRhL = fetal rhesus lung cells

PGMK = primary green monkey kidney cells

DEN-1 16007 Parent

- Recovered from serum of a patient with hemorrhagic fever and shock in Thailand in 1964
- Passaged 3X in BS-C-1 cells, 1X in LLC-MK_2 cells
- Passaged 2X in *Toxorhynchites amboinensis* mosquitos
- PDK-1

↓

PDK-43 Vaccine

DEN-2 16681 Parent

- Recovered from serum of a patient with hemorrhagic fever and shock in Thailand in 1964
- Passaged 3X in BS-C-1 cells, 1X in LLC-MK_2 cells
- Passaged 2X in *Toxorhynchites amboinensis* mosquitos
- PDK-1

↓

PDK-53 Vaccine

DEN-3 16562 Parent

- Recovered from serum of a patient with hemorrhagic fever and shock in the Philippines in 1964
- Passaged 3X in BS-C-1 cells, 1X in LLC-MK_2 cells
- Passaged 2X in *Toxorhynchites amboinensis* mosquitos
- PGMK-1

PGMK-30

DEN-3 virus grown in PGMK cells replicated to very low titer in cells (Yoksan *et al.*, 1986)

PDK FRhL-3 Vaccine

DEN-4 1036 Parent

- Recovered from serum of a patient with dengue fever in Indonesia in 1976
- Passed 4X in *Aedes aegypti* mosquitos
- PDK-1

PDK-48

The DEN-2 full-length cDNA clone was derived from the DVBID seed of DEN-2 16681 virus, which had the passage history:

Human
3X BS-C-1 cells
2X LLC-MK_2 cells
2X *T. amboinensis* mosquitos
4X C6/36 cells (*Aedes albopictus*)

Complementary DNA (cDNA) was amplified by RT/PCR directly, without further cell culture passage, from virus present in vaccine vials of the DEN-1 PDK-43, DEN-2 PDK-53, DEN-3 PGMK-30/FRhL-3, and DEN-4 PDK-48 viruses.

Stock virus seed was prepared from virus-infected cells grown in 75 or 150 cm^2 plastic tissue culture flasks. The culture medium was clarified by centrifugation for 30

min at 10,000 rpm in a Sorvall GSA rotor, bringing the final concentration of fetal bovine serum (FBS) to 10% (v/v), and then freezing the clarified virus suspension in aliquots of 0.5 - 1.0 ml at -70°C. Gradient purified DEN-2 16681 virus was prepared according to the method of Obijeski *et al.* (1976) as reported by Kinney *et al.* (1983).

Cell Lines:

Infectious virus was derived from the infectious cDNA clones by electroporation of BHK-21-15 (baby hamster kidney-21, clone 15) cells with transcribed viral RNA. Viruses were also grown in LLC-MK_2 monkey kidney cells, Vero African green monkey kidney cells, and C6/36 mosquito cells (*Aedes albopictus* C6 cells, clone 36, Igarashi (1978)). All four cell lines were grown in Eagle's minimal essential medium (MEM) supplemented with 10% (v/v) heat-inactivated (56°C for 30 min) FBS, 1.25 g/L of sodium bicarbonate, 100 units/ml of penicillin G, and 100 μg/ml of streptomycin sulfate. Confluent cell monolayers grown in plastic tissue culture flasks were infected by decanting the growth medium, permitting the virus inoculum to adsorb for 1.5 h at 37°C, and then adding MEM containing 5% FBS. For plaque titration of viruses, confluent cell monolayers in plastic 6-well trays were inoculated with 200 μl of the appropriate dilution of virus. Virus was adsorbed to the cell monolayer for 1.5 h at 37 °C. The cells were then overlaid with 3 ml of 1% (w/v) Noble agar (maintained at 40°C) in MEM lacking

phenol red pH indicator and containing 2% FBS and 0.01% (w/v) DEAE-dextran. Following incubation for 6 days at 37 °C in a 5% CO_2 atmosphere, a second 1-ml agar overlay containing 50 μg/ml of neutral red vital stain was added. Viral plaques were counted 2-5 days later.

E. coli:

The *E. coli* K-12 strains used in this project included XL1-Blue, MC-1061, SURE, JM101, and TB-1. Recombinant plasmid containing full genome-length cDNA of DEN-2 virus was successfully replicated in *E. coli* XL1-Blue, MC-1061, and TB-1. Flavivirus cDNA, particularly the gene region encoding the envelope glycoprotein, is troublesome in *E. coli*. Bacteria hosting the recombinant plasmid containing the full-length cDNA clone grew slowly and were often difficult to streak for isolation on agar plates containing selective antibiotic. Transformation efficiencies were sometimes improved somewhat by incubation of agar plates at 30°C or ambient temperature rather than at 37°C. Bacterial stocks were stored frozen at -70°C in 10% (v/v) glycerol.

Precautions for Working with RNA:

RNA is a fragile molecule that is very readily degraded by the many ubiquitous RNases present in the environment. Many of these RNases are resistant to treatment with detergents and heat, including autoclaving. All reagents and materials that contacted the viral RNA in

this project were RNase-free to avoid degradation of the viral RNA by these ubiquitous, very stable enzymes. The investigator wore tight-fitting gloves, maintained all reagents on ice, used a plastic tool to open the lids of microtubes, used individually packaged pipets, preferably plastic for aqueous solutions, disposable plasticware which is generally RNase-free before opening, and used "For RNA Only" microtubes, Gilson micropipetors (P-10, P-20, P-100, P-200, P-1000) and tips with aerosol barriers. Use of recycled glassware was avoided. Weigh boats, magnetic stirrers, and pH meters were not used. Chemicals were weighed in sterile, RNase-free disposable plastic 50-ml centrifuge tubes, and solutions were adjusted to the appropriate pH by aliquoting a small volume of the solution onto pH paper. Whenever possible, commercially prepared, guaranteed RNase-free reagents were purchased. Otherwise, newly-opened chemicals were reserved "For RNA Only". Water and stock salt solutions, except for those containing Tris, were treated overnight with 0.1% (v/v) diethylpyrocarbonate (DEPC) to inactivate RNases via alkylation and then autoclaved for 20 min. It is advisable to use the best sterile technique when working with RNA.

Extraction of Viral Genomic RNA from Virus Seed:

Virus seeds containing at least 10^6 PFU (plaque forming units)/ml of virus are ideal for providing appropriate yields of RNA. Seed with virus titer of 10^4 or lower can be problematic in terms of yielding sufficient

RNA. For these low-titer seeds it is best to pool the yields of several extracted seed aliquots.

RNA extraction involved the addition of 200 μl of cold RNA lysis buffer (4 M guanidine isothiocyanate, 25 mM sodium citrate, pH 7.0, 0.5% (w/v) sarkosyl, and 100 mM beta-mercaptoethanol), and 30 μl of 3 M sodium acetate, pH 5.2, to an empty RNase-free 1.5-ml microtube on ice. In a biosafety cabinet, 200 μl of DEN virus seed was added to the microtube and mixed vigorously for 30 sec with a mechanical mixer. The tube was centrifuged briefly to pellet the liquid; then 400 μl of cold phenol (commercially supplied by AMRESCO) equilibrated to pH 4.5 and 80 μl of cold chloroform were added. The tube was mixed vigorously for 30 sec, placed on ice for 15 min, mixed again, then centrifuged for 1 min at maximum speed in a refrigerated microcentrifuge to separate the aqueous and organic phases. The top aqueous phase containing the extracted RNA was transferred to a fresh 1.5-ml microtube on ice, 400 μl of cold isopropanol was added, and the tube was incubated for at least 1 h or overnight at -20°C. The RNA was precipitated by centrifugation for 10 min at maximum speed at 4°C. The supernatant was removed with a pipet rather than by decantation and rinsed with 500 μl of 75% (v/v) ethanol. After spinning again for 10 min, the ethanol was removed with a pipet. The tube was centrifuged again briefly and the residual liquid was removed with a micropipet. The RNA pellet was air dried briefly, resuspended in 50 μl of cold RNase-free dH_2O, and stored frozen. For seeds containing low virus titer, the

RNA pellets in 3-6 microtubes were pooled in a total volume of 50 μl.

RT/PCR Synthesis of Dengue Virus-Specific cDNA Fragments

Full-length genomic mRNA was extracted directly from 200 μl of DEN virus seed. The standard reverse transcriptase/polymerase chain reaction (RT/PCR) was performed in a 100-μl reaction solution containing 5-18 μl of the extracted viral RNA, 1 μl each of 100 μM stock solutions (stored frozen in dH_2O) of the upstream mRNA-sense primer-amplimer and downstream complementary-sense primer-amplimer, 10 μl of 10X standard PCR buffer (500 mM KCl, 100 mM Tris-HCl, pH 8.5, 15 mM $MgCl_2$ and 0.1% (w/v) gelatin), 8.0 μl of 2.5 mM dNTPS (2.5 mM each of dATP, dCTP, dGTP, and dTTP; Pharmacia-LKB), 0.5 μl of 1 M dithiothreitol (DTT), 0.5 μl of RNase inhibitor (RNasin, 40 U/μl, Boehringer-Mannheim), 0.5 μl of Taq DNA polymerase (5 U/μl, Perkin-Elmer), and 0.5 μl of RAV-2 reverse transcriptase (18 U/μl, Takara). The reaction solution was made as two components:

▸ PCR Reaction Mix:	10.0 µl	10X Standard PCR Buffer
	8.0 µl	2.5 mM dNTPs
	0.5 µl	1 M DTT
	0.5 µl	RNasin (40 U/µl)
	0.5 µl	Taq DNA Polymerase (5 U/µl)
	0.5 µl	RAV-2 RT (18 U/µl)
	60.0 µl	RNase-Free dH_2O
	80.0 µl	Reaction Mix for 1 reaction. Make more than needed for all reaction tubes. Store excess at -70°C for reuse.
▸ Template/Primer Mix:	18.0 µl	DEN-2 RNA Template
	1.0 µl	100 µM Up-Amplimer
	1.0 µl	100 µM Down-Amplimer
	20.0 µl	
▸ Reaction Solution:	80.0 µl	PCR Reaction Mix
	20.0 µl	Template/Primer Mix
	100.0 µl	In a thin-walled, 200-µl microtube.

The RT/PCR reactions in thin-wall 200-µl microtubes (Phenix Research Products) were incubated without oil overlay in a Perkin-Elmer Model 9600 thermocycler according to the following program:

50 °C for 60 min =	First strand cDNA synthesis by reverse transcriptase
94°C for 4 min	
50°C for 1 min	
72°C for 5 min	
94°C for 30 sec	30 Cycles
55°C for 30 sec	
72°C for 5 min	
Delta +10 sec/cycle	

Following completion of the RT/PCR reactions, 5-μl aliquots of each of the 100-μl reactions were analyzed by agarose gel electrophoresis. The DNA bands in the agarose gel were stained in ethidium bromide (500 ng/ml) solution and visualized on an ultraviolet light box. Since extraneous non-target cDNA bands are often amplified in addition to the target cDNA molecules, the remaining 95 μl of each RT/PCR reaction was electrophoresed in a larger, preparative agarose gel, and the target cDNA was stained briefly, excised with a razor blade, and physically extracted from the agarose slice.

High-Melt-Crush (HMC) Extraction of DNA from Agarose:

An agarose gel slice containing DNA was placed in a 1.5-ml microtube and crushed thoroughly with a spatula or pestle. The volume of the crushed agarose was brought to 400-500 μl with TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM disodium EDTA) and 400 μl of phenol (supplied by AMRESCO), pH 8, was added. The agarose suspension was mixed vigorously using a mechanical mixer, frozen, thawed and mixed, frozen, thawed and mixed, and then centrifuged for 10 min at maximum speed at 4°C. The top aqueous phase was transferred to a fresh microtube, extracted with 400 μl of phenol:chloroform:isoamyl alcohol (25:24:1) and centrifuged for 2 min. The top aqueous phase was transferred to a fresh tube and extracted with 700 μl of diethyl ether or chloroform. If chloroform was used, the top phase was again transferred to a fresh tube after a brief spin to separate phases. The DNA was precipitated

for at least 30 min at -70°C or overnight at -20°C following addition of 2.5 volumes (essentially filling the microtube) of 95% ethanol containing 300 mM ammonium acetate and 10 mM $MgCl_2$. The DNA was pelleted at 4°C by centrifugation for 20 min at maximum speed. The liquid was decanted, and the DNA pellet was rinsed with 500 μl of 75% ethanol, air-dried briefly, dissolved in 30 μl of TE buffer, and stored frozen or in the refrigerator. A 3-μl aliquot of the extracted DNA was analyzed for purity and quantity by agarose gel electrophoresis. Generally, 20-80% of the DNA loaded onto a gel can be recovered from the gel by this method.

Agarose Gels:

DNA was analyzed by electrophoresis in 1% (w/v) agarose gels run in TBE buffer (100 mM Tris-HCl, pH 8, 91 mM boric acid, and 20 mM disodium EDTA). DNA bands were visualized by staining the gel in water containing 500 ng/ml of ethidium bromide and exposure to ultraviolet light. Gels used for analyzing RNA transcripts were made with RNase-free reagents. Ethidium bromide stain was incorporated in the gel and running buffer so that the RNA bands could be visualized immediately. To obtain gel-purified DNA fragments, DNA was electrophoresed in 0.7% (w/v) agarose gels made with genetic technology grade Seakem agarose (FMC) or with biotechnology grade agarose (3:1 high resolution blend, AMRESCO).

Cloning of Dengue Virus-Specific cDNA Fragments:

Some DNA polymerases add an extra "A" nucleotide overhang at the 3'-end of synthesized DNA strands. The Taq DNA polymerase does this. To enable the cloning of DNA molecules synthesized using Taq DNA polymerase, TA-cloning vectors have been engineered (Marchuk *et al.*, 1991). These vectors generally have a single "T" overhang engineered at the 3'-terminus of EcoRV-cut, blunt-ended, linearized plasmid vector. The EcoRV site occurs within the multiple cloning site (MCS) of the plasmid. The MCS is a series of contiguous, unique restriction enzyme (RENZ) sites engineered into a vector plasmid to permit subcloning of exogenous DNA fragments following restriction with a variety of RENZs. The HMC-purified DEN cDNA amplicons were cloned into the 3900-bp pCRII (Invitrogen), the 2887-bp pT7Blue(R) (pT7Blue, Novagen), or the 3003-bp pGEM-5Zf (Promega) TA-vector plasmid. The RENZ sites available in the MCS region of these TA-vectors, as well as the RENZ sites of the MCS of the general purpose cloning plasmids, pUC18 and pUC19, used in this project are shown below.

RENZ Sites Present in the MCS of Several Cloning Vectors

pUC18	pUC19	pT7Blue	pCRII	pGEM-5Zf
		T7	SP6	T7
EcoRI	HindIII	HindIII	NsiI	ApaI
SstI	SphI	BspMI	HindIII	AatII
KpnI	PstI	SphI	KpnI	SphI
SmaI	SalI	PstI	SstI	NcoI
BamHI	XbaI	Sse8387I	BamHI	SstII
XbaI	BamHI	SalI	SpeI	**EcoRV**
SalI	SmaI	AccI	BstXI	SpeI
PstI	KpnI	HincII	EcoRI	NotI
SphI	SstI	XbaI	**EcoRV**	PstI
HindIII	EcoRI	SpeI	EcoRI	SalI
		NdeI	PstI	NdeI
		EcoRV	BstXI	SacI
		BamHI	NotI	BstXI
		AvaI	AvaI	NsiI
		SmaI	SphI	SP6
		KpnI	NsiI	
		SacI	XbaI	
		BanII	ApaI	
		EcoRI	T7	

The pUC18/19 plasmids possess identical MCS sites in reverse orientation in the plasmid backbone. Their purpose is to permit cloning of DNA in either orientation into the plasmid using the same pair of RENZs - this

reversibility was exploited in this project. The TA-vectors used here all possessed T7 and/or SP6 bacteriophage RNA promoters to enable RNA transcription from cloned DNA. These promoters were not used in this project. All of the plasmids contain the gene for ampicillin resistance. They also contained the *lac Z* portion of the *E. coli lac* operon. This permits color discrimination between bacterial colonies that receive a recombinant or a wild-type plasmid. In the presence of IPTG and X-gal, bacterial colonies that are transformed with a wild-type plasmid lacking a cDNA insert develop a blue color, whereas cells that receive a recombinant plasmid with cDNA cloned into the MCS of the plasmid are white. Agar plates contained 800 μg of IPTG and 800 μg of X-gal.

Fifty to 100 ng of HMC-purified amplicon was ligated to 50 ng of the pCRII vector using the TA-vector cloning kit supplied by Invitrogen exactly as specified by the instructions supplied with the kit. Frozen, transformation competent E. coli INVαF' cells, supplied with the Invitrogen kit and stored at -70°C, were transformed with the ligated DNA as described in the kit instructions. The transformed cells were plated on YTA_{50} agar plates (8 g of DIFCO tryptone, 5 g of DIFCO yeast extract, 5 g of NaCl, and 15 g of BACTO agar per liter of dH_2O) containing 50 μg/ml of ampicillin. Only bacterial cells transformed with the pCRII plasmid, which contains an ampicillin resistance gene, grow on this medium. The agar plates were incubated at 37°C overnight.

Similarly, cDNA was ligated to the other TA-vectors or to pUC18/19 cut with the appropriate RENZ(s). Ligations were performed at room temperature or at 12°C. *E. coli* XL1-Blue, SURE, TB-1, or MC-1061 cells were transformed by electroporation and plated on YTA_{50} plates. Electroporation was performed according to Dower *et al.* (1988) using cuvettes with a 2-cm electrode gap in a Bio-Rad Gene Pulser set at 2.5 kV voltage, 25 μF capacitance, and 200 ohms resistance. Electroporation-competent cells were prepared by growing a fresh bacterial culture to an optical density of 0.5-0.7 at 600 nm. The cells from 1.5 - 3 L of culture were pelleted by centrifugation for 10 min at 4°C and 5000 rmp in a Sorvall GSA rotor, pooled, washed twice in 1 mM Hepes buffer, and resuspended in 2 ml of 10% (v/v) sterile glycerol per L of original culture. The concentrated cells in glycerol were stored at -70°C.

Bacterial colonies were transferred to 2 ml of 2XYT-Amp_{50} broth (16 g of tryptone, 20 g of yeast extract, and 5 g of NaCl per liter of dH_2O) and incubated overnight with shaking at 300 rpm at 37°C in a floor model incubator - shaker (model Innova 4300, New Brunswick). Recombinant plasmid was extracted from these 2-ml minicultures and analyzed by agarose gel electrophoresis for the presence of cDNA insert. Recombinant plasmids are larger than wild type vector plasmid because of the cDNA insert, and they migrate more slowly than wild type plasmid in agarose gels.

All of the DEN-2 16681 virus-specific cDNA amplicons were cloned into the pCRII TA-vector. Aliquots of insert-positive miniprep plasmids were digested with the

restriction enzyme EcoRI. Since the pCRII MCS contains two EcoRI recognition sites (palindromic hexameric sequence GAATTC) on either side of the EcoRV cDNA cloning site, this RENZ cleaved the cDNA insert from the plasmid vector and cleaved any EcoRI sites that were present within the cDNA itself. The EcoRI-restricted DNA was analyzed by agarose gel electrophoresis to determine that the cloned cDNA was of appropriate size. In our experience, cloning of PCR-derived cDNA amplicons 2000 bp or smaller in size into the TA-vector is efficient. Cloning amplicons larger than 3500 bp into the TA-vector can be very difficult.

After screening, certain of the miniprep plasmids were selected for further analysis. Their corresponding bacterial minicultures were streaked for isolation on YTA_{50} plates, and an isolated colony was inoculated into 50-200 ml of YTA_{50} broth to grow up a preparative amount of recombinant plasmid. The preparative scale for the extraction of the plasmid was essentially identical to that for minipreps except for scaled up volumes.

Extraction of Plasmid DNA from Minicultures of *E. coli*:

White colonies containing recombinant plasmid were picked with a sterile toothpick and shaken overnight at 300 rpm in 2 ml of $2X\text{-}YTA_{50}$ broth. Each miniculture was decanted into a 1.5-ml microtube, and the cells were pelleted by centrifugation at 6000 rpm for 2 min. The supernatant was aspirated, and the cell pellet was resuspended gently by up/down micropipeting in 200 μl of

GTE buffer (50 mM glucose, 25 mM Tris-HCl, pH 8.0, and 25 mM disodium EDTA) and then mixed with 300 μl of lysis buffer (0.2 N NaOH, 1% (w/v) sodium dodecylsulfate (SDS)). After incubation on ice for 5 min, 300 μl of cold potassium acetate solution (3 M potassium acetate, 7 M acetic acid, pH 4.8) was added, and the solution was chilled for 5 min on ice and then centrifuged at maximum speed for 10 min at 4°C. The supernatant was poured into a fresh microtube, RNase A was added to 20 μg/ml, and the mixture was incubated at 37°C for 30 min. The sample was extracted twice with 600 μl of chloroform and centrifuged for 1 min at maximum speed at room temperature. The DNA pellet was dissolved in 32 μl of dH_2O. Eight μl of 4M NaCl and 40 μl of 13% (w/v) PEG-8000 was added, and the mixed solution was incubated for 5 min on ice. The sample was centrifuged for 15 min at maximum speed at 4°C, the liquid was aspirated with a micropipet , and the pellet was rinsed with 500 μl of 75% ethanol. The air dried pellet was dissolved in 30 μl of dH_2O and stored frozen until used.

Extraction of Plasmid DNA from Large Cultures of *E. coli*:

Preparative-scale plasmid extraction was performed by inoculating 100 ml of 2X-YTA_{50} broth with 2 ml of an overnight culture of *E. coli*. The culture was shaken overnight at 300 rpm and 37°C. The cells were pelleted by centrifugation for 10 min at 5000 rpm in a Sorvall GSA rotor and resuspended in 6 ml of cold GTE buffer. Nine ml of a freshly made solution of 0.2 N NaOH and 1% (w/v) SDS

was added. The sample was incubated for 5 min on ice, then 9 ml of cold 3 M potassium acetate solution was added. After another 5-min incubation on ice, the tube was centrifuged for 20 min at 10,000 rpm at room temperature and the supernatant was transferred to a fresh 30-ml glass tube. RNase A was added to 20 μg/ml, and the sample was incubated for 30 min at 37°C and then extracted twice with 6 ml of chloroform. Twelve ml of room-temperature isopropanol was added and the tube was centrifuged immediately for 20 min at 10,000 rpm at room temperature. The supernatant was decanted, and the DNA pellet was rinsed with 1 ml of 75% ethanol, air dried briefly, and resuspended in 480 μl of dH_2O. The DNA was precipitated by addition of 120 μl of 4 M NaCl and 600 μl of 13% PEG-8000, incubation for 5 min on ice, and centrifugation for 15 min at maximum speed at 4°C. The DNA pellet was rinsed with 500 μl of 75% ethanol, air dried briefly, rehydrated in TE buffer, and stored frozen.

Nucleotide Sequence Analysis of the Dengue cDNA Clones:

Nucleotide sequence analyses of DEN-2 16681 cDNA clones #1-#15 were performed by cloning EcoRI restriction fragments of each clone into the single-stranded bacteriophage M13mp18 or M13mp19. Since this is not the current method of choice for sequencing, the method will be described only briefly here. The procedure used for the extraction of plasmid DNA from bacterial cells was also used to extract the intracellular double-stranded replicative form (RF) DNA of M13 from bacteriophage-

infected *E. coli* JM101 cells. The RF DNA was linearized at the EcoRI site of the MCS and ligated to the DEN-2 HMC-purified EcoRI cDNA restriction fragments.

Electroporation-competent *E. coli* JM101 cells were transformed by electroporation and plated onto H-agar plates (10 g of DIFCO tryptone, 5 g of NaCl, 15 g of BACTO agar, and 1% (w/v) thiamine per liter of dH_2O) containing 800 μg each of isopropyl-β-D-galactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (BCIG or X-gal). The electroporated cells were mixed with 300 μl of a fresh logarithmic culture of JM101 cells and 3 ml of warm (51°C) top H-agar containing 9 g/L of agar and then poured onto the H-agar plates. Cells that were transfected with recombinant DNA supported replication of recombinant M13 virus, resulting in the formation of bacteriophage plaques in the JM101 cell lawn on the agar plate. The IPTG/BCIG histochemistry of the system permitted identification of white plaques containing recombinant bacteriophage into which cDNA had been ligated into the EcoRI site of the MCS, whereas wild-type nonrecombinant M13 bacteriophage produced blue plaques. Isolated plaques were picked, inoculated into 3 ml of a fresh, pre-logarithmic phase culture of JM101, and shaken at 37°C for 8-16 h. The minicultures were clarified by centrifugation in 1.5-ml microtubes, the bacteriophage particles were precipitated with PEG-8000, and the single-stranded, circular bacteriophage DNA was isolated from the virions by phenol extraction. The recombinant, circular, single-stranded bacteriophage DNA was sequenced by the dideoxynucleotide termination method. Sequencing kits can

be purchased from various commercial vendors. Radioactive ^{32}P-dCTP or ^{35}S-dCTP was incorporated into the strands synthesized in the sequencing reactions. Sequencing was accomplished with many DEN-2 virus-specific primers designed to sequence the entire genome. The sequence reactions were electro-phoresed in 6% (w/v) polyacrylamide gels, which were dried onto filter paper and overlaid with X-ray film. The DNA bands of the autoradiographs were read by the investigator, and the data was entered into a sequence project data spreadsheet. This sequencing method has been used extensively in the past (*e.g.*, Kinney *et al.*, 1986; Johnson *et al.*, 1986; Deubel *et al.*, 1986; Deubel *et al.*, 1988; Kinney *et al.*, 1989; Trent *et al.*, 1987).

Nucleotide sequencing was also performed by the current method of direct sequencing of double-stranded plasmid DNA by the dideoxynucleotide termination method using the Applied Biosystems Taq DyeDeoxy Terminator Cycle Sequencing Kit, cycle sequencing in the Model 9600 thermocycler according to the instruction manual supplied with the kit, and analyzing the DNA sequence on an ABI Model 373A DNA sequencing apparatus. Sequencing reactions in 200-μl thin-walled microtubes contained 9.5 μl of reaction mix (buffer, the four dideoxynuleotides, and Taq polymerase supplied in the kit), 7.0 μl of double or single-stranded template DNA (150 pg/bp), and 3.2 μl of 10 μM sequencing primer (32 pmol). After mixing, the reactions were placed in a Perkin-Elmer Model 9600 thermocycler, and programmed cycle sequencing was performed for 25 cycles of incubation at 96°C for 15 sec,

50°C for 15 sec, and 60°C for 4 min. Strand extension was performed at 60°C rather than 72°C because the fluorescent dye-labeled dideoxynucleotide terminators are heat sensitive. The reaction was then applied to a Centrisep gel column (Princeton Separations) to remove unincorporated dye-labeled dideoxynucleotides according to the instructions supplied with the columns. The eluted DNA was vacuum dried for 1 h using a Savant Speed Vac Concentrator and stored at -70°C. The DNA was hydrated with 5 μl of deionized formamide and 1 μl of 50 mM disodium EDTA, then heated in an aluminum block for 2 min at 90°C. A 3-μl aliquot of the denatured DNA sample was applied to one of 24 wells of a polyacrylamide-urea gel in an Applied Biosystems 373A DNA sequencer. The color-coded sequence chromatograph was read by visual inspection, and the resulting nucleotide sequence was entered into a computer-maintained sequence data spreadsheet. The sequencing kit incorporates dideoxynucleotide terminators that are each labeled with a unique fluorescent dye that permits laser detection of all four terminators in a single polyacrylamide gel lane in the Model 373 sequencer. The data was recorded in the form of colored chromatograms that are easily read by the investigator. Single-stranded recombinant M13 DNA can also be sequenced in this manner.

<u>Extraction of M13 Single-Stranded DNA for Sequencing:</u>

White bacteriophage plaques containing recombinant M13 DNA were picked with sterile toothpicks and placed into 2-ml slightly turbid (less than 0.15 A_{600}) cultures of

E. coli JM101. The cultures were shaken at 300 rmp and 37°C overnight and then clarified by centrifugation in microtubes at maximum speed for 10 min at room temperature. One ml of the supernatant was transferred to a fresh 1.5-ml microtube containing 200 μl of sterile 20% (w/v) PEG-8000 in 250 mM NaCl. The tubes were mixed by inversion, incubated for 15 min at room temperature, and centrifuged at maximum speed for 5 min at room temperature. The PEG supernatant was removed completely, and the DNA pellet was resuspended in 300 μl of TE buffer. An equal volume of pH 8-buffered phenol was added, and the solution was mixed vigorously several times during a period of 20 min at room-temperature. The tube was centrifuged for 5 min at room temperature, and the top aqueous phase was transferred to a fresh 1.5-ml microtube. After sequential extraction with phenol:chloroform:isoamyl alcohol and chloroform, the DNA was precipitated by adding 2.5 volumes of 95% ethanol containing 300 mM ammonium acetate and 10 mM $MgCl_2$ and incubating at -20°C overnight. The tube was centrifuged at maximum speed for 15 min at 4°C, and the supernatant was decanted. Following a rinse with 500 μl of 75% ethanol, the DNA was air dried briefly, resuspended in 60 μl of TE buffer, and stored at 4°C.

Primers:

Primer design was based on the sequence of DEN-2 virus, strain 16681, published by Blok *et al.* (1992), and DEN-2 virus, Jamaican strain 1409, as reported by Deubel *et al.* (1986) and Deubel *et al.* (1988).

Primers were synthesized by the Biotechnology Core Facility at the CDC in Atlanta, Georgia. We received the dried primers via mail and adjusted them to a concentration of 100 μM in dH_2O. The designations and sequences of all of the primers-amplimers used in this project are listed in Appendix A.

To amplify the 3'-end of the DEN-2 virus genome, a downstream amplimer was designed that was complementary to the published sequence of the 3' terminus of the genome. A unique XbaI restriction enzyme site was incorporated at the 5' end of this amplimer to provide a unique site to permit linearization of the recombinant plasmid containing the full-length cDNA clone at the 3' terminus of the cloned genomic cDNA. This linearization was necessary to obtain appropriately terminated DEN virus-specific run-off RNA transcripts from the cDNA clone in transcription reactions with bacteriophage T7 RNA polymerase. Linearization at this 3'-terminal XbaI site resulted in the incorporation of a 5-nucleotide TCTAG extension to the 3' terminus of the genomic mRNA transcribed from the full-length cDNA clone of DEN-2 16681 virus, and a 4-nucleotide CTAG extension to the 3' terminus of RNA transcribed from the DEN-2 PDK-53 cDNA clone. The difference between the two cDNA clones in the length of the extraneous 3'-terminal extension was due to the differently designed 3'-terminal amplimers used to obtain the 3' end genomic cDNA amplicon. Amplimer cD2-10687.XBA or cD2-10687.X2 was used to amplify and clone the 3'-terminal portion of DEN-2 16681 or PDK-53 virus, respectively.

The promoter for the bacteriophage T7 RNA polymerase was engineered at the 5' terminus of the cloned genomic cDNA by incorporating the recognition sequence of the T7 RNA polymerase into the sequence of the 5'-terminal upstream, mRNA-sense amplicon D2-SMT71 immediately preceding the 5'-terminal nucleotide of the DEN-2 viral genome. This design ensured that the T7 RNA polymerase initiated RNA transcription at the 5'-terminal nucleotide of the DEN-2 virus-specific cDNA (Milligan *et al.*, 1987).

Amplimers for PCR reactions were designed to take advantage of RENZ sites identified within the nucleotide sequence of the genome of DEN-2 16681 virus. cDNA molecules were amplified to permit ligation or splicing together of overlapping contiguous cDNA clones at shared, overlapping, unique RENZ sites (Figure 3).

Transcription of Genomic mRNA from DEN Virus-Specific Full-Length cDNA Clones:

The recombinant plasmid containing the full-length cDNA clone was prepared for RNA transcription by linearization at the unique XbaI site located at the 3' terminus of the cloned genomic cDNA. The restriction reaction containing the XbaI-restricted plasmid was extracted sequentially with phenol:chloroform:isoamyl alcohol and chloroform and then precipitated. The DNA was redissolved in 50 μl of TE buffer and digested with proteinase K at a concentration of 1 mg/ml for 1 h at 37°C to hydrolyze contaminating RNases. The sample was then extracted twice with "For RNA Only" phenol:chloroform:isoamyl alcohol buffered to pH 8,

extracted twice with chloroform to remove traces of phenol, and precipitated by adding one-tenth volume of RNase-free 3 M sodium acetate, pH 5.2, and 2.5 volumes of ethanol and incubating for at least 1 h at -70°C or overnight at -20°C.

DEN-2 virus-specific genomic RNA was transcribed from the linearized cDNA template using a commercial T7 transcription kit (Ampliscribe T7 transcription kit, Epicentre Technologies). Transcription reactions were performed for 2 h at 37°C in RNase-free 1.5-ml microtubes in 20-μl reactions containing 100-1000 ng of linearized DNA template, 7.5 mM each of CTP, GTP, and UTP, 0.75 mM ATP, 2.7 mM m^7GpppA cap analog, 6.7 mM DTT, 2.0 μl of a 10X concentration of a proprietary buffer supplied with the commercial kit, and 2.0 μl of the proprietary Ampliscribe enzyme solution supplied with the kit. Reaction solutions were used directly and without further treatment to transfect BHK-21 cells.

Transfection of BHK-21 Cells with Genomic RNA Transcripts:

BHK-21 clone 15 cells were transfected with RNA transcripts by electroporation (Liljeström *et al.*, 1991). Fresh cultures of BHK-21 cells were grown to 90% confluency, rinsed twice with cold RNase-free phosphate buffered saline (PBS), and released from the plastic by incubation with 3 ml of commercial trypsin-EDTA solution (GIBCO-BRL). The cells were pelleted by low-speed centrifugation at 1200 rpm for 5 min in a Beckman GPKR centrifuge. The cells were washed twice with cold PBS,

resuspended in cold PBS and kept on ice. The cells were counted using a hemacytometer and microscope, and the cell concentration was adjusted to 10^7 cells/ml. One-half ml of the washed, adjusted cells were mixed with each transcription reaction solution in 1.5-ml microtubes on ice. The mixture was transferred to a cold electroporation cuvette with 0.2-cm electrode gap, which was placed in the cuvette holder of the Bio-Rad Gene Pulser. The cells were shocked twice using settings of 1.5 kV voltage, 25 μFD of capacitance, and resistance set to infinity. The shocked cells were incubated for 10 min at room temperature and then added to 75 cm^2 tissue flasks containing 20 ml of MEM containing 10% FBS. Transfected cell cultures were incubated at 37°C for 5-8 days until CPE was evident in the cell monolayer and/or expression of DEN virus-specific antigens was identified in an aliquot of the cell monolayer scraped from the flask using DEN virus-specific mouse hyperimmune ascitic fluid or monoclonal antibodies in indirect immunofluorescence tests.

RESULTS

Replication of DEN-2 16681 Virus:

DEN-2 16681 virus replicates to high titer in cell culture. The CDC virus seed used in this study contained 2.0 X 10^7 plaque forming units (PFU)/ml. This titer was determined by plaque titration of the seed virus in monolayer cultures of Vero cells. This seed titered 1.3 X 10^4 PFU/ml in LLC-MK_2 cells. A growth curve for this virus

was determined in C6/36 *Aedes albopictus* cell culture (Figure 4). This level of replication is quite high for a flavivirus. The DEN-2 16681 virus is eminently suitable to serve as the parent to an infectious cDNA clone of DEN virus.

The DEN-2 PDK-53 vaccine virus, taken directly from a vaccine vial obtained from Mahidol University, contained 3.4×10^4 PFU/ml of virus, as titrated in Vero cell monolayers, and 1.5×10^4 PFU/ml as titrated in LLC-MK_2 cell monolayers.

RT/PCR Amplification and Cloning of DEN-2 16681 cDNA:

The entire genome of DEN-2 virus, parental strain 16681, was amplified from genomic RNA in the form of 5 cDNA clones of various sizes (T7-F1, F2, F3, F4, and F5). PCR amplification with 5 sets of upstream and downstream amplimers yielded the predicted amplicon sizes in PCR reactions. Figure 5 shows the migration of these cDNA fragments in agarose gels.

Recombinant plasmids, obtained by ligating the cDNA amplicons into the pCRII TA-vector, were extracted from minicultures derived from transformed *E. coli* XL1-Blue colonies. Uncut plasmids were screened for the presence of cDNA insert by comparing their mobility in agarose gels with the mobility of uncut wild-type pCRII vector plasmid. Selected plasmids were then restricted with the restriction enzyme EcoRI to confirm the size of the inserted cDNA fragment. EcoRI digests of F2-Sal, Sal-F2,

and F3 plasmids derived from independent transformed bacterial colonies are shown in Figure 6.

The following 15 DEN-2 16681 virus-specific cDNA clones, shown schematically in Figure 7, were selected for nucleotide sequence analysis:

Clone	RT/PCR Amplicon	
1	F1	- A8
2	F1	- A21
3	F1	- A25
4	F1	- A26
5	F2-Sal	- AA2-4
6	F2-Sal	- AA2-8
7	Sal-F2	- AA3-3
8	Sal-F2	- AA3-4
9	F3	- AA4-4
10	F3	- AA4-6
11	F4	- 10
12	F4	- 12
13	F5	- AA6-1
14	F5	- AA6-2
15	F5	- AA6-4

RT/PCR Amplification and Cloning of DEN-2 PDK-53 cDNA:

The entire genome of DEN-2 virus, vaccine strain PDK-53, was amplified from genomic RNA in the form of 23 cDNA clones of various sizes. Even though the PDK-53 vaccine contained only about 10^4 PFU/ml of virus, we were able to routinely amplify cDNA from RNA that was extracted directly from this seed virus. To accomplish this, we routinely use the "extended PCR method", incorporating the Taq extender reagent (Stratagene) in the PCR reactions. We had previously shown that the Taq extender

significantly enhanced yields of large molecular weight amplicons in the PCR amplification of the nonstructural genes of the flavivirus, St. Louis encephalitis virus (Figure 8A). For extended PCR reactions, reaction mixtures were made as for standard PCR reactions, but the standard PCR buffer was replaced with the Taq extender buffer and 1 unit of AmpliTaq DNA polymerase (Perkin-Elmer) and 1 unit of the Taq extender enzyme per kbp of expected amplicon size was included in the reaction. Figure 8B shows the correct agarose gel migration of large cDNA amplicons F1 (containing the T7 RNA polymerase promoter at the 5' end of the mRNA-sense strand of the amplicon), F2, and F3 obtained by PCR amplification using DEN-2 PDK-53 viral genomic RNA as template. The standard PCR reaction also worked for a number of DEN-2 PDK-53 amplifications.

The PDK-53 PCR products were cloned into the pGEM-5Zf TA-vector (Promega) or the pT7Blue(R) TA-vector (Novagen). Although we seemed to have the best cloning efficiency of PCR amplicons in the pCRII TA-vector, the other vector kits were less expensive and worked well. The cloning efficiency of PCR products into the TA-vector decreased rapidly as amplicon size increased beyond 2000 bp.

The following 23 DEN-2 PDK-53 virus-specific cDNA clones were selected for nucleotide sequence analysis:

CLONE	RT/PCR AMPLICON	Expected Amplicon Length	Up-Amplimer	Down-Amplimer
1	F-5	1552-bp	D2-SMT71	cD2-1510
2	F1-7	"		"
3	F1-9	"		"
4	F1-75A	"		"
5	F1-79B	"		"
6	F2-14	3355-bp	D2-1261	cD2-4615
7	F2-16B	"		"
8	F3-33	2676-bp	D2-4257	cD2-6932
19	F3-3C	"		"
10	F4-9	2373-bp	D2-6493	cD2-8865
11	F4.9-22	2937-bp	D2-6493	cD2-9429
12	F4.9-53	"		"
13	F4.5-1	1897-bp	D2-8440	cD2-10337
14	F4.5-2	"		"
15	F4.5-6	"		"
16	F4.5-7	"		"
17	F5-72	1914-bp	D2-8773	cD2-10687.X2
18	F5-77	"		"
19	F5-78	"		"
20	F3.5-4	1375-bp	D2-6046	cD2-7420
21	F3.5-6	"		"
22	F3.5-19	"		"
23	F3-3K	2676-bp	D2-4257	cD2-6932

Nucleotide Sequence Analyses of DEN-2 16681 cDNA Clones:

EcoRI fragments of the 15 DEN-2 16681 virus-specific cDNA clones were subcloned into the single-stranded bacteriophage M13mp18 or M13mp19 for sequencing. Sequencing of the entire viral genome was performed manually using radioisotopic labeling and exposure,

development, and reading of autoradiographs. The data was read from the films and entered by hand into a sequence data spreadsheet.

The locations of observed cDNA artifacts or "errors" dictated the splicing strategy of subclones to construct the full genome-length clone. If the nucleotide at a particular position of one cDNA clone differed from the nucleotides at that same position in 2 or more independent clones, then the nucleotide in the first clone was deemed to be an error. If only 2 cDNA clones were sequenced for a given region of the genome and they differed in sequence at a particular position, then if one of the cDNA clones agreed with the sequence data of Blok *et al.* (1992), then the clone containing the nucleotide that was in agreement with the latter investigators was deemed to be correct. The approximate locations of the cDNA errors identified in the 16681 clones are illustrated in Figure 9.

The full genome-length cDNA clone of DEN-2 16681 virus was first constructed in pUC19. Unfortunately, RNA transcribed from this clone was not infectious. When over 90% of the full-length cDNA in the clone was resequenced, it was determined that several mutations had occurred during splicing and cloning manipulations of the subclones in *E. coli*. One of these mutations was a base deletion in the NS4B gene. This deletion would cause a frameshift of the amino acid sequence, resulting in ribosomal translation of a nonsense polypeptide downstream of the mutation point. This fatal deletion, by itself, would explain the noninfectious nature of the RNA transcribed from the first full-length clone in pUC19.

The final, correct cDNA subclones (F1-E, F2-E, F3/4/5-F) that were incorporated into the full-length, successfully-infectious clone of 16681 virus were reanalyzed by direct sequencing of the double-stranded plasmid DNA via the thermocycling method using the Taq DyeDeoxy Terminator Cycle Sequencing Kit. Sequence analysis was performed using the automated 373A DNA sequencing machine. The color-coded sequence chromatograms were read by the investigator and the data was entered manually into a computer-based spreadsheet.

We independently confirmed the sequence of the 5'-terminal 32 nucleotides of the DEN-2 16681 viral genome. A 5'-end RNA-cDNA hybrid molecule, made with primer cD2-996 and reverse transcriptase, was 3'-tailed with dCTP and annealed to dGTP-tailed, PstI-cut M13mp19 RF DNA. One of the resulting M13 clones had a cDNA run-off product containing the 5'-terminal end of the genome. The 5'-end sequence was identical to that published for DEN-2 1409 (Deubel *et al.*, 1988) and DEN-2 16681 (Blok *et al.*, 1992). We have not independently confirmed the sequence of the 3'-terminal 36 nucleotides of DEN-2 16681 virus or the 5'- or 3'-terminal nucleotides of DEN-2 PDK-53 virus.

We sequenced uncloned, PCR-derived amplicon cDNA fragments directly for the following regions of the DEN-2 16681 viral genome: nucleotides 70-260, 330-870, 890-1690, 1890-3720, 3770-4050, 4080-4320, and the 3'-terminal 9990-10686. Unlike the sequencing of cloned DNA, direct analysis of PCR amplicons provides sequence information for the majority population of amplified cDNA molecules,

and therefore for the majority population of template RNA molecules.

We observed very early in the project that the nucleotide sequence of DEN-2 16681 virus that we determined at the CDC laboratory differed significantly from the sequence of DEN-2 16681 virus as published by Blok *et al.* (1992). Our nucleotide sequence differed from that published by Blok *et al.* (1992) at 60 nucleotide positions, which were located throughout the genome. Amino acid substitutions were encoded by 26 of these nucleotide differences. The approximate genomic locations of the nucleotide differences are illustrated in the schematic diagram in Figure 10. The exact nucleotide positions of the discrepancies are shown in Figure 11.

Nucleotide Sequence Analyses of DEN-2 PDK-53 cDNA Clones:

The DEN-2 PDK-53 virus-specific cDNA clones were analyzed by direct sequencing of the double-stranded plasmid DNA by the thermocycling method using the Taq DyeDeoxy Terminator Cycle Sequencing Kit. The 3'-end sequence from nucleotide position 10290-10686 was also determined by direct sequencing of PCR-derived amplicon cDNA. Sequence analysis was performed using the automated 373A DNA sequencing machine. The color-coded sequence chromatograms were read by the investigator and the data was entered manually into a computer-based spreadsheet. The approximate locations of the cDNA errors identified in the PDK-53 cDNA clones are illustrated in Figure 12.

Our determination of the nucleotide sequence of DEN-2 PDK-53 virus differed significantly from the PDK-53 genomic sequence published by Blok *et al.*. (1992). The latter investigators reported a total of 53 nucleotide differences that encoded 27 amino acid mutations between the nucleotide sequences of the genome of DEN-2 16681 virus and that of its vaccine derivative, PDK-53 virus. They reported the following nonsilent mutations: 1 in the capsid, 2 in prM, 1 in M, 3 in E, 3 in NS1, 3 in NS2A, 2 in NS2B, 3 in NS3, 3 in NS4A, 3 in NS4B, and 3 in NS5. We detected only 8 nucleotide mutations between the genomes of these two virus strains. One mutation occurred in the 5'-NC region of the genome, while 7 nucleotide mutations, 4 of which encoded amino acid substitutions, occurred in the coding region of the genome as shown in Figure 13 and the following table.

Table: Summary of nucleotide differences between the genomes of DEN-2 16681 virus and its vaccine derivative virus, strain PDK-53.

		Genome				
Genome			Nucleotide		Amino Acid	
Position		Gene	16681	PDK-53	16681	PDK-53
57[a]		5'-NC	C	T	-	-
524[a]		prM-29	A	T	Asp	Val
2055[a]		E-373	C	T	Phe	Phe
2579[a]		NS1-53	G	A	Gly	Asp
4018		NS2A-151	C	T	Leu	Phe
5547		NS3-342	T	C	Arg	Arg
6599[a]		NS4A-75	G	C	Gly	Ala
8571[a]		NS5-334	C	T	Val	Val

[a] 16681 vs. PDK-53 difference agrees with Blok *et al.* (1992)

The few nucleotide positions where our data and those of Blok *et al.* (1992) agreed, in terms of sequence differences between the 16681 and PDK-53 viral genomes, were distributed throughout the genome. The entire genome of DEN-2 16681 virus was cloned and sequenced before we received the PDK-53 vaccine virus at our laboratory.

Except for the 3'-terminal cDNA clones #17-#19, every PDK-53 virus-specific cDNA clone constructed in our laboratory contained at least one nucleotide position of 16681/PDK-53 sequence difference confirmed by both ourselves and Blok *et al.* (1992). Therefore, our PDK-53 virus-specific cDNA clones did not result from contamination of PDK-53-specific PCR reactions with 16681 virus-specific cDNA template. Our PDK-53 virus-specific cDNA clones, which also contained the many sequence discrepancies between our data and those of Blok *et al.* (1992), encoded the nucleotide sequence from the 5' terminus to nucleotide position 10337 of the genome of PDK-53 virus. The 3'-terminal 387 nucleotides (10337-10723) of DEN-2 PDK-53 virus were identical to those of the parental 16681 virus. Since none of the PDK-53 virus-specific cDNA clones covering this region of the genome contained a point of confirmed 16681/PDK-53 sequence difference, we repeated the PCR amplification of the 3' terminus of the PDK-53 virus genome. This was done to ensure that the 3'-terminal cDNA clones #17-#19 did not result from PCR reactions contaminated by 16681 virus-specific DNA template. The PCR reaction components were pipetted in a room in which DEN cloning had not been performed previously, using new micropipetors, newly opened pipet tips with aerosol barrier, and freshly made stock reagents. Direct sequencing of the resulting double-stranded PCR cDNA amplicon confirmed that the 3'-387 nucleotides of DEN-2 PDK-53 virus was indeed identical to the 3' terminus of the 16681 parent.

The finalized nucleotide sequence of DEN-2 virus, strain 16681, including the nucleotide and amino acid mutations identified for DEN-2 PDK-53 virus, is shown in Figure 14.

Construction of DEN-16681 Full-Length Clone in pUC19:

For the construction of the full genome-length cDNA clone of DEN-2 16681 virus, 5 of the sequence-characterized PCR-amplified cDNA subclones were selected for splicing. However, clone #5 contained a cDNA "error" that was not readily spliced out with the existing clones. This error, which was a C-to-T mutation at nucleotide position 1730 and encoded a nonsilent Thr-to-Ile amino acid substitution at E-265, was incorporated into the F2 construct. The intermediate F2 construct was the result of splicing the F2-Sal clone (#5) SphI/HpaI fragment to the Sal-F2 clone (#7) HpaI/KpnI fragment in the MCS of plasmid pUC18 (Figure 15). To correct the error, a new PCR amplicon was made using primers D2-1261 and cD2-2955. Resulting clones in the TA-vector were sequenced, and the correct SphI/HpaI fragment of a new clone was substituted for the faulty SphI/HpaI fragment of the original F2 construct (Figure 16). The corrected F2 clone was designated F2-C.

The relevant cDNA clones of DEN-2 16681 virus were spliced together via a series of intermediate ligation products in the MCS of pUC18 to yield F1/3/4/5, which contained all of the genome except for the SphI-KpnI 1380-4493 region present in clone F2-C. Multiple attempts to

ligate the F2-C SphI/KpnI cDNA fragment into F1/3/4/5 in pUC18 failed. The cDNA insert of F1/3/4/5-pUC18 was then transferred to the MCS of pUC19, resulting in F1/3/4/5-pUC19. This operation simply reversed the orientation of the cDNA insert within the context of the pUC plasmid. Ligation of SphI/KpnI-cut F1/3/4/5-pUC19 and F2-C SphI/KpnI insert readily yielded transformants in *E. coli* Xl1-Blue that contained the full-length cDNA clone F1/2/3/4/5-pUC19, which was designated pD2/IC-20. The detailed splicing procedures for pD2/IC-20 are illustrated in Figure 17. The orientation-specific cloning of the full genome-length cDNA in pUC19 rather than pUC18 is diagrammed in Figure 18.

The full genome-length cDNA of DEN-2 16681 virus was cloned into the MCS of pUC19. Apparent full genome-length viral mRNA was transcribed from linearized pD2/IC-20. This transcribed product failed to yield infectious virus following electroporation of BHK-21 cells. Most of the cDNA in the pD2/IC-20 clone was resequenced, and several cloning artifacts, including a fatal single-nucleotide deletion, were identified. Original subunit intermediate cDNA constructs in pUC18 were resequenced to confirm that they possessed the correct sequence and corrected where necessary. The corrected primary cDNA clones F1, F2-C, and F3/4/5 were then ligated into the low-copy plasmid pBR322, rather than the high copy-number pUC18 plasmid. It was envisioned that the cDNA would be more stable in a slower-replicating plasmid in *E. coli*.

To enable more straightforward cloning into pBR322, the MCS of pUC19 was spliced into the pBR322 plasmid

(Figure 19). This resulted in plasmids pBRUC-138 and pBRUC-139 containing the pUC MCS in both orientations within the pBR322 plasmid backbone. The SphI site was removed from both pBRUC plasmids by cutting with SphI, blunt ending of the cut ends using T4 DNA polymerase, and then ligating the ends back together. This was necessary for the construction of the full-length cDNA clone because SphI is one of the cDNA restriction/splicing sites for the clone.

The F3/4/5-F cDNA clone of DEN-2 16681 virus, which had been verified by sequence analysis, was cloned into pBRUC-139 ($SphI^-$) (Figure 20). Following this ligation, the F1-E and F2-C cDNA clone fragments were also moved into the pBR322 backbone to construct the full genome-length cDNA clone, pD2/IC-30P (Figure 20). This recombinant plasmid was replicated successfully in both TB-1 and MC-1061 strains of *E. coli*.

Construction of DEN-2 PDK-53 Infectious cDNA Clone:

The full-length infectious clone of DEN-2 16681 virus was used in the construction of the infectious clone for PDK-53 virus. Since the 3'-noncoding regions of the genomes of both viruses are identical, and the amino acid sequences of the translated precursor polyproteins encoded by genome nucleotide positions 6646-10269 are identical in both viruses, the infectious clone of PDK-53 virus was constructed using the 16681 3'-end cDNA from the NheI site at nucleotide position 6646 to the 3' terminus of the genome (Figure 21). After correcting a cDNA error in the

PDK-53 F3-3C subunit clone, this fragment and the F2-16B cDNA fragment were ligated into the infectious clone backbone to construct the DEN-2 PDK-53 virus-specific full-length cDNA clone, pD2/IC-130V (Figure 21).

Transcription of Viral mRNA from DEN-2 Infectious cDNA Clones:

Viral genomic RNA extracted from gradient-purified virions was analyzed by nondenaturing RNA agarose gel electrophoresis to observe the level of RNA degradation and the limits of detectability by ethidium bromide staining. Figure 22 shows an agarose gel electropherogram for 22-383 ng of viral genomic RNA obtained from purified preparations of wild-type DEN-2 16681 virus and wild-type Venezuelan equine encephalitis (VEE) virus, strain Trinidad donkey. Although degradation of the RNA is visible as a spectrum of smaller molecular weight nucleic acid (smear in Figure 22), definite full-genome length RNA bands are clearly visible. This smear of nucleic acid is probably also due, in part, to multiple conformations of the single-stranded RNA molecules which migrate through the gel at different rates. The relative gel migration of the single-stranded RNA does not correlate directly with the sizes of the double-stranded molecular weight marker DNA bands (MW, Figure 22); the VEE and DEN-2 viral genomes are 11,447 and 10,723 nucleotides in length, respectively. BHK-21 and C6/36 cells were transfected successfully by electroporation with 2000, 500, 100, 10, 1, and 0.1 ng of viral genomic RNA extracted from purified VEE or DEN-2

16681 virus, as indicated by development of CPE, expression of viral proteins detected by indirect immunofluorescence tests using virus-specific antibody, and/or by plaque titration of infectious virus from the transfected-cell culture medium. RNA quantities of 1 ng or less were essentially undetectable in the ethidium bromide-stained agarose gel system we used. Therefore, authentic RNA transcripts derived from full genome-length cDNA and visualized in agarose gel electropherograms of transcription reactions should be infectious for BHK-21 cells by electroporation.

Investigators previously constructed an infectious cDNA clone for VEE virus as reported by Kinney *et al.* (1989). RNA transcription reaction conditions that yielded high quantity and quality of infectious mRNA transcripts from the pVE/IC-92 infectious clone of VEE virus failed in multiple attempts to transcribe RNA from the pD2/IC-20 clone of DEN-2 16681 virus. Figure 23 shows an agarose gel electropherogram that demonstrates successful transcription of RNA from the VEE clone, but not pD2/IC-20.

In an attempt to improve RNA transcription from the DEN-2 clone, commercial transcription kits were purchased. The Megascript transcription kit supplied by Ambion also failed to transcribe RNA from the DEN clone. However, the Ampliscribe kit obtained from Epicentre Technologies enabled efficient transcrip-tion of RNA from the DEN-2 clone (Figure 24).

The success of the Ampliscribe kit apparently was due to the high concentration of ribonucleotides and a very

high, but proprietary, concentration of T7 RNA polymerase. The RNA transcribed from pD2/IC-20 was not infectious. However, viral mRNA transcribed from DEN-2 16681 clone pD2/2-IC30P and PDK-53 clone pD2/IC-130V was infectious (Figure 25).

Viral mRNA transcripts from both replicates of pD2/IC-30P (A and D) and pD2/IC-130V (F and J) were infectious, producing viable infectious virus in electroporated BHK-21 cells. Figure 26 shows RNA transcripts from pD2/IC-20, pD2/IC-30P, and pD2/IC-130V.

Construction of DEN-2 16681/PDK-53 Chimeric cDNA Clones:

Several chimeric full-length cDNA clones were derived from the pD2/IC-30P and pD2/IC-130V clones. All clones were constructed in the pBRUC-139 derivative of the pBR322 plasmid vector. *E. coli* strains XL1-Blue, MC-1061, and TB-1 were successfully transformed with ligated recombinant plasmids containing full genome-length cDNA. Viable virus was derived from all of the indicated clones. The evolutionary tree for the chimeric viruses is diagrammed in Figure 27.

Details concerning the splicing strategies for the chimeric clones are shown in Figure 28. Appropriate cDNA fragments were cut and ligated together at the internal SalI, SphI, KpnI, and NheI sites as well as at the 5'-SstI and 3'-XbaI sites.

Viable prototype and chimeric viruses were derived from each of the clones indicated in Figure 28 by electroporation of BHK-21 cells with viral genome-length

mRNA transcribed from linearized plasmids. Seed stocks of these viruses were prepared by centrifuge-clarification of the cell culture medium, adjustment of the FBS concentration to 10%, and freezing of seed aliquots at -70°C. Virus concentrations were determined by plaque titration of the virus seeds in monolayer cultures of Vero cells. The results of these virus titrations are shown in the following table.

Table. Plaque titration of DEN-2 16681 and PDK-53 stock seed viruses and chimeric viruses recovered from BHK-21 cells transfected with infectious clone-derived viral mRNA transcripts.

Virus	(PFU/ml)	Genotype[a]
DEN-2 16681	8.0×10^7	c D F G L R G V
DEN-2 PDK-53	5.1×10^3	t V . D F . A .
D2/IC-30P-A	3.6×10^5	. . . . .
D2/IC-30P-A2	1.7×10^5	. . . . .
D2/IC-130V-F	4.0×10^5	t V . D F . A .
D2/IC-130V-J	2.2×10^5	t V . D F . A .
D2/IC-130V2-1	2.8×10^5	t V . . . . A .
D2/IC-130V2-7	8.8×10^4	t V . . . . A .
D2/IC-31-12	2.1×10^5	t V . . . . .
D2/IC-31-15	3.2×10^5	t V . . . . .
D2/IC-32-A	1.4×10^6	. . . D F . . .
D2/IC-32-G	1.2×10^6	. . . D F . . .
D2/IC-33-C	9.6×10^4	. . . . . A .
D2/IC-33-P	1.9×10^5	. . . . . A .
D2/IC-321-L	1.1×10^6	t V . D F . . .
D2/IC-321-N	7.6 X 105	t V . D F . . .
D2/IC-323-B	7.2×10^5	. . . D F . A .
D2/IC-323-I	8.8×10^5	. . . D F . A .
D2/IC-31-57-5	2.4×10^5	t . . . . .
D2/IC31-524-D	3.2×10^4	c V . . . . .

[a] Genotype is designated in small case for the virus-specific 5'-noncoding nucleotide and in upper case single-letter amino acid abbreviation for amino acids encoded by virus -specific nucleotide mutations. Dots represent nucleotide or amino acid sequence identity with DEN-2 16681 virus.

To establish the validity of the clone-derived chimeric viruses, relevant genomic cDNA fragments were amplified directly from seed viruses by PCR and spot-sequenced. The results are shown in Figure 29. This validation process is ongoing. Except for D2/IC-31-524 virus, appropriate cDNA insert regions in chimeric viruses have been confirmed by sequence analysis. Except for D2/IC-30P, D2/IC-130V, and D2/IC-31-57, which have been fully confirmed, clone-derived chimeric viruses have yet to be spot-sequenced in a recipient clone-derived cDNA region to definitely establish the chimeric nature of the virus. The recipient clone is the recombinant plasmid backbone into which a cDNA fragment, the insert fragment, from a heterologous donor clone is spliced. Where duplicate clone-derived viruses were obtained, both viruses of a given genotype were spot-sequenced, and both gave the same result, which is shown in Figure 29.

Submission of pD2/IC-30P and pD2/IC-130V to ATCC:

Patent deposits of the full genome-length cDNA clones of DEN-2 16681 and PDK-53 viruses were submitted to the American Type Culture Collection (ATCC), Rockville, Maryland, U.S.A. Both pD2/IC-30P-A and pD2/IC-130V-F were grown overnight in *E. coli* TB-1 cells. Six cryogenic vials containing 1 ml each of frozen cell culture in 10% glycerol were submitted by dry ice shipment. Prior to shipment, plasmid was extracted from a 1 ml aliquot of each virus-specific culture. The recombinant full-length

plasmid was recovered from the cells as shown in Figure 30.

The pD2/IC-30P-A deposit with the ATCC was assigned accession number ATCC 69826, and the pD2/IC-130V-F deposit with the ATCC was assigned accession number ATCC 69825. Date of deposit was May 25, 1995.

Construction of Chimeric DEN-2/1, -2/3, and -2/4 Infectious Clones:

We contemplate deriving chimeric DEN-2/1, DEN-2/3, and DEN-2/4 viruses from recombinant full genome-length cDNA clones containing the genetic background of DEN-2 PDK-53 virus and the prM and E genes of the DEN-1, DEN-3, and DEN-4 candidate vaccine viruses, respectively. To accomplish this, the prM and E genes of the vaccine viruses were amplified by PCR. Because our laboratory has been establishing a sequence database to analyze the molecular epidemiology of several flaviviruses, including all of the serotypes of dengue virus, the primers used for cDNA amplification in the PCR were readily available at our laboratory. The amplified cDNA molecules were sequenced directly, thus providing the sequence of the population of virions in the virus seed. The amplified cDNA amplicons for the DEN-1, DEN-3, and DEN-4 vaccine viruses have all been cloned into the pGEM-5Zf TA-vector. The cloned cDNA has not been analyzed by sequencing, since it will be necessary to rederive the cDNA amplicons by PCR to incorporate appropriate RENZ cleavage sites within the amplicon for splicing into the full-length cDNA backbone of DEN-2 PDK-53 virus. The partial nucleotide sequences

of the genomes of the DEN-1, DEN-3, and DEN-4 vaccine viruses were aligned with the DEN-2 PDK-53 sequence. All four sequences are aligned with the nucleotide sequence of DEN-2 16681 virus and its deduced amino acid sequence in Figure 31. The deduced amino acid sequences of the DEN viruses are aligned in Figure 32.

It is readily evident from the aligned nucleotide sequence data that useful restriction enzyme sites in the DEN-2 virus-specific cDNA are not conserved in the DEN-1, DEN-3, and DEN-4 viruses. Therefore, splicing sites must be engineered into the cDNA to enable the splicing of heterotypic DEN-1, DEN-3, and DEN-4 prM and E genes into the DEN-2 backbone. It is not yet clear precisely how the nonstructural proteins of flaviviruses interact with the structural proteins during intracellular maturation of the virus. Furthermore, the interaction of the capsid protein with the genomic mRNA molecule in the nucleocapsid of the virion has not been defined. However, coexpression of the E and prM proteins has been more successful than expression of E alone in expression systems *in vitro*. The DEN-2 nonstructural proteins are involved in all virus-specific intracellular polyprotein processing and replication of viral mRNA, and the predominant portion of the mRNA genome interacting with the capsid protein is presumably, but not necessarily, DEN-2 virus-specific. For these reasons, our strategy is to splice in the prM and E genes of DEN-1, DEN-3, and DEN-4 viruses very precisely, while maintaining the DEN-2 context of the bracketing capsid and NS1 protein regions.

The strategies for creating XhoI and XbaI splice sites at the 5' end of the prM gene and near the 3' end of the E gene are illustrated in detail in Figures 33 and 34, respectively. Briefly, mutagenic primers containing the appropriate RENZ site are utilized in PCR reactions to synthesize new cDNA for the prM and E genes of all four viruses. A DEN-2 PDK-53 virus-specific cDNA cassette plasmid, designated pD2V-CAS12, containing the genome region from the 5' terminus through nucleotide position 4696 is constructed via intermediate plasmid constructs pF1-Xho and pF2-Xba as illustrated in Figures 35 and 36. The XhoI/XbaI cDNA fragments cut directly from DEN-1, DEN-3, and DEN-4 virus-specific amplicons synthesized by PCR using the mutagenic primers are ligated into the pD2V-CAS12 cassette plasmid to create subclone chimeras. The SstI/KpnI fragment of the resulting pD1V-CAS12, pD3V-CAS12, and pD4V-CAS12 cassettes are moved into pD2/IC-130V restricted with SstI/KpnI to create the chimeric full genome-length cDNA clones (Figure 36).

Discussion:

Infectious cDNA clones permit the directed engineering of viral genomes. Depending on their viability in terms of ability to replicate in cell culture, infectious clone-derived viruses can be modified by incorporating point mutations, multiple mutations, deletions, gene regions of related or heterologous viruses, or nonviral genes. Infectious cDNA clones have been developed for many RNA viruses, including flaviviruses DEN-4 (Lai et al., 1991), yellow fever (Rice

et al., 1989), Kunjin (Khromykh and Westaway, 1994), Japanese encephalitis (Sumiyoshi et al., 1992), and TBE (unpublished data). We describe herein the development of infectious cDNA clones for DEN-2 16681 virus and its candidate vaccine derivative, strain PDK-53. We also describe the construction of chimeric viruses, incorporating the prM and E genes of candidate DEN-1, DEN-3, and DEN-4 vaccine viruses within the genetic background of the DEN-2 PDK-53 vaccine virus.

Although the candidate vaccine viruses developed at Mahidol University are currently the best live DEN virus vaccine candidates in terms of immunogenicity and safety in adult humans, the DEN-1, DEN-3, and DEN-4 vaccine viruses replicate poorly in cell culture and possess low infectivity in humans, requiring up to 2000-fold more PFU of virus to infect and immunize humans than is needed for the DEN-2 PDK-53 vaccine virus. The low infectivities of these viruses have significant implications for vaccine production in cell culture, potentially decreased immunogenic efficacy, and more rapid inactivation under conditions of a poorly maintained cold chain in tropical countries where dengue viruses are endemic.

The purpose of engineering chimeric DEN vaccine viruses is to enhance the replicative ability and immunogenicity of the DEN-1, DEN-3, and DEN-4 vaccine viruses. A primary assumption has been that the attenuated DEN-2 PDK-53 vaccine virus replicates to appropriate levels in cell culture. In fact, it does appear that the genome of DEN-2 PDK-53 virus is eminently suited to serve as the genetic backbone for chimeric

viruses containing the prM and E genes of DEN-1, DEN-3, and DEN-4 vaccine viruses. We have recently completed growth curves for DEN-2 16681 virus, DEN-2 PDK-53 virus, and their infectious clone derivative viruses in LLC-MK_2 cells.

The viruses were titrated in Vero cell monolayers. These data are shown in the following table:

Virus	Maximum Titer (PFU/ml)	Maximum Titer at Day
DEN-2 16681	2.6×10^8	10
D2/IC-30P-A	1.7×10^7	8
D2/IC-30P-A2	6.6×10^7	7
DEN-2 PDK-53	3.8×10^7	9
D2/IC-130V-F	2.9×10^7	7
D2/IC-130V-J	1.7×10^7	7

The DEN-2 PDK-53 virus and its infectious clone derivative viruses grow to approximately 10^7 PFU/ml in LLC-MK_2 cells, about as well as the DEN-2 16681 virus.

A second assumption is that the chimeric DEN viruses will be viable and the DEN-2 PDK-53 virus-specific replication machinery will significantly increase replication of the chimeric viruses in cell culture and increase their infectivity and immunogenicity in humans relative to the wild-type vaccine viruses. The high degree of conservation of amino acid sequences among the polyproteins of the four DEN viruses should ensure that the chimeric viruses will be viable. The level of

replication attained by the chimeric DEN viruses is determined empirically, as was determined for the DEN-2 PDK-53 infectious clone derivative virus.

Bray et al. (1991) constructed chimeric DEN-4/1 and DEN-4/2 viruses that appeared to appropriately express DEN-1 and DEN-2 structural protein antigens in the genetic background of DEN-4 virus. These investigators spliced much of the 5'-noncoding region, and the capsid, prM and E genes of DEN-1 or DEN-2 virus into the full-length cDNA clone of DEN-4 virus. The near 3'-terminal splice site they chose in the E gene is very close to that proposed by us in our project. These chimeric viruses replicated very slowly relative to the wild-type viruses. The authors attributed this slow replication to possible suboptimal gene expression, assembly, and/or maturation due to incompatibility of heterotypic genes or RNA packaging in the nucleocapsid. Another possibility is that cDNA errors may have been incorporated into their constructs. In contrast, Pletnev et al. (1993) engineered chimeric viruses between DEN-4 virus and tick-borne encephalitis (TBE) virus, which is a very distant flavivirus relative of DEN viruses. Thus, DEN virus chimeras may be derived that are viable.

A third assumption is that our chimeric DEN viruses will express the appropriate structural protein antigens of DEN-1, DEN-3, and DEN-4 viruses, and that vaccinees will respond with development of appropriate serum titers of DEN-1, DEN-3, and DEN-4 neutralizing antibodies following immunization with the chimeric viruses. We describe the insertion of the prM and E genes of DEN-1,

DEN-3, and DEN-4 viruses into the DEN-2 clone. Thr-to-Ser amino acid substitutions near the amino terminus of the prM protein in DEN-2, DEN-2/1, DEN-2/3, and DEN-2/4 viruses resulting from mutagenesis to create the XhoI site of the cassettes should be conservative in nature and affect the phenotype of derived viruses minimally, if at all. Alternatively, a unique MluI site (ACGCGT) could be created via a single, silent A-to-G point mutation at nucleotide position 453 in the DEN-2 clone. The MluI site immediately preceding the T7 promoter could easily be eliminated by cutting the clone with MluI, blunt-ending, and religation. The clone-derived DEN-2 and chimeric viruses would then have the prM amino-terminal sequence "FHLTTR."

The carboxyl-terminal 24 amino acids of the E glycoprotein of all of the infectious clone-derived viruses will be those of the DEN-2 PDK-53 virus. Therefore, the E protein of all of the chimeric viruses will have amino acid mutations in this region. Yet, the carboxyl-terminal 39 amino acids of the DEN virus E protein comprise membrane-spanning, transmembrane domains. In all enveloped viruses, the transmembrane domains of the integral viral proteins of related viruses are quite variable in amino acid sequence. It has often been noted that the important conserved feature of amino acids in this domain lies in their hydrophobic, "lipid-loving" nature rather than in the absolute sequence. Creation of a MroI site (TCCGGA) or a unique AgeI site (ACCGGT) at nucleotide positions 2281-2286 in the DEN-2 clone would

result in amino acids "SG" or "TG", respectively, at positions E-449 and E-450 in the clone-derived viruses.

The E protein of all flaviviruses share a similar gross tertiary structure that is indicated by the absolute conservation of the 6 Cys residues in the prM protein and in the 12 Cys residues in the ectodomain (the region located on environment side of the viral lipid envelope) of the E protein of DEN, Japanese encephalitis, West Nile, Murray Valley encephalitis, St. Louis encephalitis, Kunjin, yellow fever, TBE, Langat, and Powasson flaviviruses (data not shown). Cys residues are involved in intrachain Cys-Cys disulfide bonds that determine the overall structure of the protein. We fully expect the DEN-2/1, DEN-2/3, and DEN-2/4 chimeric viruses to be viable and to replicate more efficiently than the wild-type DEN-1, DEN-3, and DEN-4 vaccine viruses, respectively. Furthermore, chimeric recombinants involving the genetic backbone of one flavivirus and the structural genes of a variety of different flaviviruses may also be viable, as has been demonstrated for DEN-4/TBE virus recombinants (Pictnev et al., 1993). Such recombinant viruses offer the potential opportunity to engineer chimeric vaccine viruses for a number of flavivirus-associated diseases within the genetic background of a single flavivirus. The X-ray crystallographic structure of the E glycoprotein of TBE flavivirus has recently been published (Rey et al., 1995). This development has significant implications for the future design of flavivirus molecular vaccines.

A fourth assumption is that the chimeric DEN viruses will retain the attenuated phenotype of the wild-type DEN-1, DEN-3, and DEN-4 vaccine viruses, despite enhanced replicative efficacy provided by the more efficient nonstructural genes and 5' and 3' noncoding regions of the DEN-2 PDK-53 virus. This presupposes that DEN-2 PDK-53 virus has attenuating mutations in the noncoding regions or in the nonstructural genes and/or that attenuating mutations occur in the prM/E region of the genomes of DEN-1, DEN-3, and DEN-4 viruses. Mutations in essentially any region of the viral genome may be capable of attenuating a virulent virus. This has been demonstrated for a number of viruses including polio virus, VEE virus, and Theiler's virus. Noncoding as well as protein coding regions may be involved in attenuation. Attenuating mutations in the envelope proteins of enveloped viruses are common (Barrett et al., 1990).

The nucleotide mutations in DEN-2 PDK-53 virus at genome nucleotide positions 57 (5'-noncoding region), 524 (prM), 2579 (NS1), 4018 (NS2A), and 6599 (NS4A) may be involved in attenuation of the virus. Unless the prM amino acid mutation is the only mutation affecting virulence of the virus, the DEN-2 PDK-53 genetic background, within which the structural genes from heterologous viruses will be expressed, does itself possess genotypic markers of attenuation. We can determine the genetic loci involved in the attenuation of the DEN-2 PDK-53 virus by analyzing DEN-2 16681/PDK-53 recombinant viruses derived from chimeric 16681/PDK-53

full-length clones. The E gene of DEN-2 PDK-53 virus contains no attenuating mutations.

Although investigators have sequenced the structural genes of numerous DEN-3 virus strains (*e.g.*, Lanciotti et al., 1994), none have sequenced the DEN-3 16562 virus, parent to the DEN-3 PCMK-30/FRhL-3 vaccine virus. After determining the sequences of the prM and E genes of this virus, we can establish if any amino acid mutations have occurred within these genes in the DEN-3 vaccine virus.

By comparison, nucleotide sequence information for the parental DEN-1 and DEN-4 viruses have been determined (unpublished data (parental DEN-1 virus); Lanciotti et al., submitted for publication (parental DEN-4 virus)). The nucleotide sequences of the E gene of DEN-4 1036 virus and both prM and E genes of DEN-1 16007 virus have been determined. The following amino acid mutations were identified:

Virus type	E Protein Amino Acid Position	Amino Acid: Parent Strain	Amino Acid: Vaccine Strain
DEN-1	E-130	Val	Ala
	E-203	Glu	Lys
	E-204	Arg	Lys
	E-225	Ser	Leu
	E-384	Ala	Glu
	E-477	Met	Val
DEN-4	E-345	Glu	Lys
	E-364	Val	Ala

There were six amino acid mutations in the E protein of DEN-1 16007 PDK-13 virus and 2 mutations in that of DEN-4 1036 PDK-48 virus. There were no amino acid substitutions in the prM protein of the DEN-1 vaccine virus. Glu-to-Lys and Lys-to-Glu amino acid substitutions, as occur at DEN-1 E-203 and DEN-4 E-345, are common motifs in sequence comparisons between parent viruses and their vaccine derivatives. It is likely that the heterologous prM/E cDNA inserts in recombinant full-length cDNA clones will transport genetic loci of attenuation into the chimeric DEN-2/1, DEN-2/3, and DEN-2/4 virus derivatives. The optimum scenario for the chimeric viruses involves increased replication ability in the presence of genetic loci of attenuation in the heterologous DEN-1, DEN-3, and

DEN-4 structural gene inserts within the genetic background of the DEN-2 PDK-53 virus.

Nucleotide sequence analysis of expressed genes is essential. The error rate in the original RT/PCR derived cDNA clones of DEN-2 16681 virus was 8.2×10^{-4}, that is 1 cDNA error for every 1227 nucleotides of cloned, sequenced cDNA. In a previous sequencing project involving VEE virus and employing classical, non-PCR cDNA synthesis methodology, the error rate was calculated to be 3.9×10^{-4} or 1 error for every 2543 nucleotides of cloned, sequenced cDNA. These errors are due to nucleotide incorporation errors by reverse transcriptase during first strand cDNA synthesis and perhaps to the cloning of individual variants within the original population of virions. Unlike many DNA polymerases, RNA polymerases and reverse transcriptase have no editing function. Incorrect nucleotides incorporated during strand elongation are not detected or removed before continuing. The Taq DNA polymerase is also known to incorporate errors into PCR amplicons. Thus, at least 4-8 cDNA "errors" can be expected to occur in 10 kb of cloned cDNA. We have observed the incorporation of spurious in-frame termination codons (TAA, TAG, TGA) in cDNA clones derived from both VEE and DEN viruses. Premature termination of amino acid translation would result in a truncated protein and would undoubtedly be a lethal mutation for a candidate infectious clone. Much of the utility of genes expressed *in vitro* is compromised when those genes are not characterized by sequence analysis. If cDNA errors occur in candidate infectious cDNA clones, it may be difficult

to determine if phenotypic effects of directed mutations are due to the engineered mutation, to cDNA errors, or to synergistic action or compensation between errors and engineered mutations.

Wiktor et al. (1984) reported that two cDNA errors caused spurious amino acid substitutions in rabies virus glycoprotein expressed in recombinant vaccinia virus and resulted in expression of non-authentic rabies glycoprotein. After sequence analysis and correction of the cDNA, expression of authentic rabies glycoprotein was obtained. A faulty cDNA clone may behave as expected in one circumstantial context, yet behave very inappropriately and be highly misleading in a different context. A faulty structural gene cDNA clone of the virulent VEE Trinidad donkey (TRD) virus that was expressed in recombinant vaccinia virus was essentially authentic by monoclonal antibody analysis of expressed VEE virus-specific proteins and by protection of immunized mice from challenge with virulent VEE virus (Kinney et al., 1988a; Kinney et al., 1988b). However, incorporation of this cDNA clone into an infectious cDNA clone of VEE virus completely abrogated the virulence of the clone-derived virus, whereas the corrected cDNA fragment resulted in derivation of virulent virus (Kinney et al., 1993).

Although Lai et al. (1991) originally derived their infectious clone of DEN-4 virus from sequence characterized subunit cDNA clones (Zhao at el., 1986; Mackow et al., 1987), the original full-length clone was not infectious (Lai et al., 1991). While these

investigators indicated that they sequenced both strands of much of the cloned genomic cDNA, they did not indicate that they sequenced more than a single clone for a given cDNA region. Nucleotides encoding cDNA errors will be confirmed on both cDNA strands, but will not be identified as errors unless the sequences of two or more independent cDNA clones covering the same region of the genome are sequenced. The functional full-length clone of DEN-4 virus was obtained by repeated splicing of large new cDNA fragments into the full-length clone until a functional clone was obtained. The authors did not indicate that the newly cloned regions were characterized by nucleotide sequence analysis (Lai et al., 1991). It is probable that the slowed replication of the DEN-4/1 and DEN-4/2 chimeric viruses relative to wild-type viruses reported by Bray et al. (1991) is due to the presence of cDNA artifacts within the full-length cDNA clone. The critical importance of accurate nucleotide sequence characterization of genes expressed *in vitro*, particularly when those genes are expressed in the form of infectious cDNA clones, is still not widely appreciated by many in the molecular biology field.

Although putative nucleotide sequences for the genomes of DEN-2 16681 and DEN-2 PDK-53 viruses have been reported in the literature (Blok et al., 1992), our sequence results indicate that the published data is highly flawed. Blok et al. (1992) reported 53 nucleotide mutations between the two viruses; we determined only 8 mutations. We analyzed at least two independent cDNA clones for regions covering the entire genomes of both

viruses. The DEN-16681 sequencing project was completed prior to receiving the DEN-2 PDK-53 virus in our laboratory, and the nucleotide sequence of the PDK-53 virus was determined from cDNA amplified directly from virus present in vaccine vials.

There are now only two classes of infectious clones developed for vaccine flaviviruses that have themselves been administered to humans: the infectious clone of yellow fever virus, vaccine strain 17D (Rice et al., 1989; Hahn et al., 1987; Rice et al., 1985), and the DEN-1, DEN-2, DEN-3, and DEN-4 vaccine derivative infectious clones described herein. Both classes of infectious clones have the important advantage of being derived from vaccine viruses that have been tested for efficacy and safety in humans. The yellow fever 17D virus vaccine has long been one of the most effective human vaccines developed; immunization with this virus provides lifelong immunity. In the case of DEN virus, it is essential that vaccines provide immunity against infection by all four serotypes of the virus. DEN-1, DEN-2, DEN-3, and DEN-4 vaccine viruses have been developed at Mahidol University, Bangkok, Thailand. All four vaccine viruses have been tested in humans and have been demonstrated to be immunogenic and safe for human adults.

Replicating vaccines in the form of live, attenuated viruses offer distinct advantages in terms of immunogenic efficacy due to replicative amplification of viral antigens (antigenic mass) in the vaccinees and replication in appropriate target tissues. Inactivated or subunit antigens usually suffer from a lack of sufficient

antigenic mass and subsequent failure to stimulate an effective immune response. Expression of proteins in recombinant vaccinia virus, which replicates primarily at the site of inoculation, may provide protection against parenteral challenge with virulent virus, but may not protect against an aerosol challenge. This was demonstrated for VEE virus when it was shown that recombinant vaccinia virus expressing the structural proteins of VEE virus protected mice from intraperitoneal challenge, but not intranasal challenge, with virulent VEE virus (Kinney et al., 1988b). Immunization with the live, attenuated VEE TC-83 vaccine virus, on the other hand, provided immunity against both parenteral challenge (immunity provided by circulating serum IgG antibody) and intranasal challenge (mucosal, IgA-base immunity) with virulent VEE virus. Furthermore, the level of immunity, as measured by titers of VEE virus-specific neutralizing antibody, were considerably higher in TC-83 virus-immunized mice and horses (the natural epidemic host for VEE virus) than in animals immunized with recombinant vaccinia/VEE virus (Kinney et al., 1988b; Bowen et al., 1992). Similar results have been reported for vaccinia/influenza A virus recombinants in rodents (Smith et al., 1986). Furthermore, a replicating vaccine virus provides the appropriate T-cell epitopes to stimulate cell-mediated immunity as well as humoral immunity. T-cell epitopes may be lacking in subunit vaccines. In short, vaccination with a safe live, attenuated vaccine virus provides the optimal immunization of a natural infection in terms of the type and level of immunity

elicited and the repertoire of viral antigens involved in generating the immune response.

To use the DEN viruses described herein as vaccine candidates, it is necessary to rederive the viruses by transfection of a cell line, such as primary dog kidney, certified for human use under conditions of good laboratory practice and management to ensure the avoidance of potential adventitious agents that might be present in uncertified cell lines. Although the cDNA-derived viruses originate from candidate vaccine viruses that have undergone testing in humans, they require recertification by analysis for possible *in vitro* phenotypic markers of attenuation and by safety testing in small animals and probably nonhuman primates. All investigative studies involving the pathogenesis of DEN virus are hampered by the unavailability of a suitable animal model. Certain *in vitro* characteristics are apparently associated with attenuation of DEN viruses, but the only definitive test is vaccine trial in human volunteers. Vaccine trails would presumably follow those of the original wild-type vaccine viruses developed at Mahidol University. The protocol includes titration of the individual vaccine virus candidates in adult human volunteers to determine the minimal infectious/immunogenic dose for each virus. This is followed by immunization trials with different bivalent and trivalent combinations of vaccine virus. The final test is the quadravalent vaccine composed of appropriate doses of all four vaccine viruses. If the preliminary trials are successful, larger trials are scheduled, and the vaccine viruses are tested in children,

who are the primary target for vaccine delivery.

We describe herein a preferred method to develop an infectious cDNA clone for a flavivirus. Optimally, a wild-type vaccine virus serves as the template for the clone construction. Large cDNA fragments are amplified from the genomic mRNA by PCR using virus-specific primers and directly cloned into a TA-vector or into the MCS of a low-copy number plasmid following restriction of the amplicon cDNA. The low-copy pBRUC-139 vector contains the MCS of pUC19 to permit convenient cloning of cDNA using a variety of RENZ sites. Other low-copy plasmids are available. The bacteriophage T7 or SP6 promoter is usually engineered into the 5'-terminal mRNA-sense amplimer, and a unique RENZ site for linearization of the recombinant plasmid containing the full-length cDNA must be engineered into the 3-terminal complementary (negative)-sense amplimer. Exhaustive nucleotide analysis of the cDNA clones is desirable.

REFERENCES

Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. SMith, and K. Struhl. 1989. **Current Protocols in Molecular Biology**. (2 volumes). Greene Publishing Associates and Wiley-Interscience, John Wiley and Sons, New York.

Barrett, A.D.T., L.M. Dunster, T.N. Ledger, C.A. Gibson, B.K. Sil, P.G. Sanders, E.A. Gould, and P.D. Minor. 1990. Evidence that the envelope protein of flaviviruses encodes determinants of attenuation and virulence. In: **Vaccines 90**. (Eds: F. Brown, R.M. Chanock, H.S. Ginsberg, and R.A. Lerner) Cold Spring Harbor Laboratory, New York, 101-104.

Bhamarapravati, N., S. Yoksan, T. Chayaniyayothin, S. Angsubphakorn, and A. Bunyaratvej. 1987. Immunization with a live attenuated dengue-2-virus candidate vaccine (16681-PDK 53): clinical, immunological, and biological responses in adult volunteers. **Bull. World Health Org.** 65:189-195.

Blok, J., S.M. McWilliam, H.C. Butler, A.J. Gibbs, G. Weiller, B.L. Herring, A.C. Hemsley, J.G. Aaskov, S. Yoksan, and N. Bhamarapravati. **Virology** 187:573-590.

Bowen, R.A., W.A. Short, C.B. Cropp, J.H. Mathews, J.T. Roehrig, R.M. Kinney, J.C. DeMartini, and D.W. Trent. 1992. Protection of horses immunized with recombinant vaccinia- Venezuelan equine encephalitis vaccine. **Vaccine Research** 1:111-121.

Brandt, W.E. 1988. Current approaches to the development of dengue vaccines and related aspects of the molecular biology of flaviviruses. **J. Infect. Dis.** 157:1105-1111.

Brandt, W.E. 1990. Development of dengue and Japanese encephalitis vaccines. **J. Infect. Dis.** 162:577-583.

Bray, M., and C.-J. Lai. 1991. Construction of intertypic chimeric dengue viruses by substitution of structural protein genes. **Proc. Natl. Acad. Sci. USA** 88:10342-10346.

Coia, G., M.D. Parker, G. Speight, M.E. Byrne, and E.G. Westaway. 1988. Nucleotide and complete amino acid sequences of Kunjin virus: definitive gene order and characteristics of the virus-specifiec proteins. **J. Gen. Virol.** 69:1-21.

Cox *et al.* 1993. Bovine Herpesvirus 1: Immune Responses in Mice and Cattle Injected with Plasmid DNA. **J. Virol.** 67:5664.

Deubel, V., R.M. Kinney, and D.W. Trent. 1986. Nucleotide sequence and deduced amino acid sequence of the structural proteins of dengue type 2 virus, Jamaica genotype. **Virology** 155:365-377.

Deubel, V., R.M. Kinney, and D.W. Trent. 1988. Nucleotide sequence and deduced amino acid sequence of the nonstructural proteins of dengue type 2 virus, Jamaica genotype: comparative analysis of the full-length genome. **Virology** 165:234-244.

Dower, W.J., J.F. Miller, and C.W. Ragsdale. 1988. High efficiency transformation of *E. coli* by high voltage electroporation. **Nucl. Acids Res.** 16:6127-6145.

Gubler, D.J. 1988. Dengue. In **The Arboviruse: Epidemiology and Ecology**, vol. II, T.P. Monath (ed.), pp. 223-260, CRC Press, Inc., Boca Raton, Florida.

Hahn, S.H., J.M. Dalrymple, J.H. Strauss, and C.M. Rice. 1987. Comparison of the virulent Asibi strain of yellow fever virus with the 17D vaccine strain derived from it. **Proc. Natl. Acad. Sci. USA** 84:2019-2023.

Halstead, S.B., S. Udomsakdi, P. Simasthien, P. Singharaj, P. Sukhavachana, and A. Nisalak. 1970. Observations related to pathogenesis of dengue hemorrhagic fever. I. Experience with classification of dengue viruses. **Yale J. Biol. Med.** 42:261-275.

Halstead, S.B., and P. Simasthien. 1970. Observations related to the pathogenesis of dengue hemorrhagic fever. II. Antigenic and biologic properties of dengue viruses and their association with disease response in the host. **Yale J. Biol. Med.** 42:276-292.

Halstead, S.B. 1984. Dengue haemorrhagic fever - a public health problem and a field for research. **Bull. World Health Org.** 58:1-21.

Halstead, S.B. 1984. Selective primary health care: strategies for control of disease in the developing world. XI. Dengue. **Rev. Infect. Dis.** 5:25264.

Halstead, S.B. 1988. Pathogenesis of dengue: challenges to molecular biology. **Science** 239:476-481.

Igarashi, A. 1978. Isolation of a Singh's Aedes albopictus cell clone sensitive to dengue and chikungunya viruses. **J. Gen. Virol.** 40:531-544.

Karabatsos, N. 1985. **International Catalogue of Arboviruses, Including Certain Other Viruses of Vertebrates**. American Society of Tropical Medicine and Hygiene, San Antonio, Texas.

Kinney, R.M., D.W. Trent, and J.K. France. 1983. Comparative immunological and biochemical analyses of viruses in the Venezuelan equine encephalitis complex. **J. Gen. Virol.** 64:135-147.

Kinney, R.M., J.J. Esposito, B.J.B. Johnson, J.T. Roehrig, J.H. Mathews, A.D.T. Barrett, and D.W. Trent. 1988a. Recombinant vaccinia virus/Venezuelan equine encephalitis (VEE) virus expresses VEE structural proteins. **J. Gen. Virol.** 69:3005-3013.

Kinney, R.M., J.J. Esposito, J.H. Mathews, B.J.B. Johnson, J.T. Roehrig, A.D.T. Barrett, and D.W. Trent. 1988b. Recombinant vaccinia virus/Venezuelan equine encephalitis (VEE) virus protects mice from peripheral VEE virus challenge. **J. Virol.** 62:4697-4702.

Kinney, R.M., G.-J. Chang, K.R. Tsuchiya, J.M. Sneider, J.T. Roehrig, T.M. Woodward, and D.W. Trent. 1993. Attenuation of Venezuelan equine encephalitis virus strain TC-83 is encoded by the 5'-noncoding region and the E2 envelope glycoprotein. **J. Virol.** 67:1269-1277.

Khromykh, A.A., and E.G. Westaway. 1994. Completion of Kunjin virus RNA sequence and recovery of an infectious RNA transcribed from stably cloned full-length cDNA. **J. Virol.** 68.4580-4588.

Lai, C.-J., B. Zhao, H. Hori, and M. Bray. 1991. Infectius RNA transcribed from stably cloned full-length cDNA of dengue type 4 virus. **Proc. Natl. Acad. Sci. USA** 88:5139-5143.

Lanciotti, R.S., J.G. Lewis, D.J. Gubler, and D.W. Trent. 1994. Molecular evolution and epidemiology of dengue-3 viruses. **J. Gen. Virol.** 75:65-75.

Lanciotti, R.S., D.J. Gubler, and D.W. Trent. (submitted). Molecular evolution and phylogeny of dengue-4 viruses. **J. Gen. Virol.** (submitted).

Mackow, E., Y. Makino, B. Zhao, Y.-M. Zhang, L. Markoff, A. Buckler-White, M. Guiler, R. Chanock, and C.-J. Lai. 1987. The nucleotide sequence of dengue type 4 virus: analysis of genes coding for nonstructural proteins. **Virology** 159:217-228.

Marchuk, D., D. Mitchell, A. Saulino, and F.S. Collins. 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. **Nucl. Acids Res.** 19:1154

Milligan, J.F., D.R. Groebe, G.W. Witherell, and O.C. Uhlenbeck. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. **Nucl. Acids. Res.** 15:8783-8798.

Obijeski, J.F., D.H.L. Bishop, F.A. Murphy, and E.L. Palmer. 1976. Structural proteins of La Crosse virus. **J. Virol.** 19:985-997.

Pletnev, A.G., M. Bray, and C.-J. Lai. 1993. Chimeric tick-borne encephalitis and dengue type 4 viruses: effects of mutations on neurovirulence in mice. **J. Virol.** 67:4956-49063.

Rey, F.A., F.X. Heinz, C. Mandl, C. Kunz, and S.C. Harrison. 1995. The envelope glycoprotein from tick-brone encephalitis virus at 2 A resolution. **Nature** 375:291-298.

Rice, C.M., E.M. Lenches, S.R. Eddy, S.J. Shin, R.L. Sheets, and J.H. Strauss. 1985. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. **Science** 229:726-735.

Rice, C.M., R. Levis, J.H. Strauss, and H.V. Huang. 1987. Production of infectious RNA transcripts from Sindbis virus cDNA clones: mapping of lethal mutations, rescue of a temperature-sensitive marker, and *in vitro* mutagenesis to generate defined mutants. **J. Virol.** 61:3809-3819.

Rice, C.M., A. Grakoui, R. Galler, and T.J. Chambers. 1989. Transcription of infectious yellow fever RNA from full-length cDNA templates produced by *in vitro* ligation. **New Biol.** 1:285-296.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. **Molecular Cloning: A Laboratory Manual**. Second Edition (3 volumes). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sedegah *et al*. 1994. Protection against malaria by immunization with plasmid DNA encoding circumsporozoite protein. **Proc. Natl. Acad. Sci. USA** 91:9519.

Smith, G.L., Jr., J.R. Bennik, J.W. yewdell, P.A. Small, Jr., B.R. Murphy, and B. Moss. 1986. Vaccinia virus recombinants expressing influenza virus genes, p. 375-389. **In Options for the control of influenza.** AP.P. Kendal and P.A. Patriarca (ed.), Alan R. Liss, Inc., New York.

Speight, G., G. Coia, M.D. Parker, and E.G. Westaway. 1988. Gene mapping and positive identification of the non-structural proteins NS2A, NS2B, NS3, NS4B and NS5 of the flavivirus Kunjin and their cleavage sites. 1988. **J. Gen. Virol.** 69:23-34.

Speight, G., and G. Westaway. 1989. Positive identification of NS4A, the last of the hypothetical nonstructural proteins of flaviviruses. **Virology** 170:299-301.

Sumiyoshi, H., C.H. Hoke, and D.W. Trent. 1992. Infectious Japanese encephalitis virus RNA can be synthesized from *in vitro*-ligated cDNA templates. **J. Virol.** 66:5425-5431.

Trent, D.W., R.M. Kinney, B.J.B. Johnson, A.V. Vorndam, J.A. Grant, V. Deubel, C.M. Rice, and C. Hahn. 1987. Partial nucleotide sequence of St. Louis encephalitis virus RNA: structural proteins, NS1, ns2a, and ns2b. **Virology** 156:293-304.

Ulmer *et al*. 1993. Heterologous Protection Against Influenza by Injection of DNA encoding a Viral Protein. **Science** 259:1745.

Westaway, E.G., M.A. Brinton, S. Ya. Gaidamovich, M.C. Horzinek, A. Igarashi, L. Kääriänen, D.K. Lvov, J.S. Porterfield, P.K. Russell, and D. W. Trent. 1985. Flaviviridae. **Intervirol.** 24:183-192.

Wiktor, T.J., R.I. MacFarlan, K.J. Reagan, B. Dietzshold, P.J. Curtis, W.H. Wunner, M.-P. Kieny, R. Lathe, J.-P. Lecocq, M. Mackett, B. Moss, and H. Koprowski. 1984. Protection from rabies by a vaccinia virus recombinant containing the rabies virus glycoprotein gene. **Proc. Natl. Acad. Sci. USA** 81:7194-7198.

Xiang *et al.* 1994. Vaccination with a Plasmid Vector Carrying the Rabies Virus Glycoprotein Gene Induces Protective Immunity against Rabies Virus. **Virology** 199:132.

Yoksan, S., N. Bhamarapravati, and S.B. Halstead. 1986. Dengue virus vaccine development: study on biological markers of uncloned dengue 1-4 viruses serially passaged in primary kidney cells. In **Arbovirus Research in Australia**, Proceedings of the Fourth Symposium. T.D. St. George, B.H. Kay, and J. Blok (eds.), pp 35-38. CSIRO/QIMR, Brisbane, Australia.

Zhao, B., E. Mackow, A. Buckler-White, L. Markoff, R.M. Chanock, C.-J. Lai, and Y. Makino. 1986. Cloning of full-length dengue Type 4 viral DNA sequences: analysis of genes coding for structural proteins. **Virology** 155:77-88.

APPENDIX A

PRIMERS DESIGNED FOR DEN-2 CLONING/SEQUENCING PROJECT:

SEQ. ID NO:	PRIMER	MER/SENSE	SEQUENCE
3	pUC/M13-P5	25/+	5'-CCCAGTCACGACGTTGTAAAACGAC-3'
4	pUC/M13-P5B	27/+	5'-GGATGTGCTGCAAGGCGATTAAGTTGG-3'
5	pUC/M13-P3	25/+	5'-TGAGCGGATAACAATTTCACACAGG-3'
6	pUC/M13-P3B	27/-	5'-GGCTTTACACTTTATGCTTCCGGCTCG-3'
7	D2-1-ECO.T7	75/+	5'-GCGGATATTG/GAATTC/TCTAGA/ AATTTAATACGACTCACTATA/ AGTTGTTAGTCTACGTGGACCGACAAAGACAG-3' (5'-Fill /EcoRI /XbaI/T7 Promoter/ 5'-end of DEN-2)
8	D2-SMT71	77/+	5'-CCAGT/GAATTC/GAGCTC/ACGCGT/ AAATTTAATACGACTCACTATA/ AGTTGTTAGTCTACGTGGACCGACAAAGACAG-3' (5'-Fill/EcoRI/SstI/MluI/T7 Promoter/ 5'-end of DEN-2)
9	D2-1	24/+	5'-AGTTGTTAGTCTACGTGGACCGAC-3'
10	D2-28	34/+	5'-GACAGATTCTTTGAGGGAGCTGAGCTCAACGTAG-3'
11	D2-134	28/+	5'-TCAATATGCTGAAACGCGAGAGAAACCG-3'
12	cD2-250	26/-	5'-GGGATTGTTAGGAAACGAAGGAACGC-3'
13	D2-274	32/+	5'-CCACCAACAGCAGGGATACTGAAAAGATGGGG-3'
14	cD2-378	25/-	5'-TGCAGATCTGCGTCTCCTATTCAAG-3'
15	D2-528	25/+	5'-CGTGAACATGTGTACCCTCATGGCC-3'
16	cD2-616	26/-	5'-TTGCACCAACAGTCAATGTCTTCAGG-3'
17	D2-616	25/+	5'-ACCAGAAGACATAGATTGTTGGTGC-3'
18	cD2-618	25/-	5'-GCACCAACAGTCTATGTCTTCTGGC-3'
19	cD2-771	25/-	5'-ATGTTTCCAGGCCCCTTCTGATGAC-3'
20	D2-847	25/+	5'-GCAGCAATCCTGGCATACACCATAG-3'
21	D2-996	27/+	5'-GGTTGACATAGTCTTAGAACATGGAAG-3'
22	cD2-996	27/-	5'-CTTCCATGTTCTAAGACTATGTCAACC-3'

SEQ. ID NO:	PRIMER	MER/SENSE	SEQUENCE
23	D2-1005	35/+	5'-GTCTTAGAACATGGAAGTTGTGTGACGACGATGGC-3'
24	D2-1141	25/+	5'-ACAACAGAATCTCGCTGCCCAACAC-3'
25	D2-1211	25/+	5'-GCAAACACTCCATGGTAGACAGAGG-3'
26	cD2-1211	25/-	5'-CCTCTGTCTACCATGGAGTGTTTGC-3'
27	cD2-1227	27/-	5'-CCACATCCATTTCCCCATCCTCTGTCT-3'
28	D2-1261	30/+	5'-GGAAAGGGAGGCATTGTGACCTGTGCTATG-3'
29	D2-1416	28/+	5'-GGAAATCAAAATAACACCACAGAGTTCC-3'
30	cD2-1503	34/-	5'-CTGCAGCAACACCATCTCATTGAAGTCGAGGCCC-3'
31	D2-1510	25/+	5'-GACTTCAATGAGATGGTGCTGCTGC-3'
32	cD2-1510	25/+	5'-GCAGCAGCACCATCTCATTGAAGTC-3'
33	D2-1546	28/+	5'-AAGCTTGGCTGGTGCACAGGCAATGGTT-3'
34	cD2-1567	27/-	5'-TGGTAACGGCAGGTCTAGGAACCATTG-3'
35	D2-1777	23/+	5'-GGACATCTCAAGTGCAGGCTGAG-3'
36	cD2-1777	23/+	5'-CTCAGCCTGCACTTGAGATGTCC-3'
37	D2-1863	27/+	5'-GAAGGAAATAGCAGAAACACAACATGG-3'
38	cD2-1888	33/-	5'-CCCTTCATATTGTACTCTGATAACTATTGTTCC-3'
39	D2-2047	32/+	5'-CCTCCATTCGGAGACAGCTACATCATCATAGG-3'
40	cD2-2047	32/-	5'-CCTATGATGATGTAGCTGTCTCCGAATGGAGG-3'
41	D2-2170	29/+	5'-ATGGCCATTTTAGGTGACACAGCCTGGGA-3'
42	cD2-2200	27/ -	5'-TGTAAACACTCCTCCCAGGGATCCAAA-3'
43	D2-2308	29/+	5'-CTCATAGGAGTCATTATCACATGGATAGG-3'
44	cD2-2504	35/-	5'-GGGGATTCTGGTTGGAACTTATATTGTTCTGTCC-3'
45	cD2-2622	30/-	5'-TGATTCAATTCTGGTGTTATTTGTTTCCAC-3'
46	D2-2702	25/+	5'-AAGGAATCATGCAGGCAGGAAAACG-3'
47	cD2-2864	22/-	5'-ACTTCCAGCGAGTTCCAAGCTC-3' (A A)
48	D2-2992	25/+	5'-AACAGAGCCGTCCATGCCGATATGG-3'
49	cD2-3105	22/-	5'-TCCATTGCTCCAAAGGGTGTGT-3' (G)
50	D2-3236	25/+	5'-AGCTTGAGATGGACTTTGATTTCTG-3'

SEQ. ID NO:	PRIMER	MER/SENSE	SEQUENCE
51	cD2-3410	22/-	5'-GGTCTGATTTCCATCCCGTACC-3'
52	D2-3621	23/+	5'-GTCCTTTAGAGACCTGGGAAGAG-3'
53	cD2-3739	25/-	5'-GTTTTCTCAAGAGTAGTCCAGCTGC-3' (C)
54	D2-3905	25/+	5'-ATCAATTGGCAGTGACTATCATGGC-3'
55	cD2-4002	25/-	5'-TGTTAAGAGCAGTGGAGAAACGGAC-3' (A, G)
56	cD2-4060	25/-	5'-GATTGAGACCTTTGATCGTCAACGC-3'
57	D2-4214	25/+	5'-TGACAGGACCATTAGTGGCTGGAGG-3'
58	D2-4257	34/+	5'-CGTGCTCACTGGACGATCGGCCGATTTGGAACTG-3'
59	cD2-4323	24/-	5'-GGGCTGCTTCCTGATATTTCTGCC-3' (C)
60	D2-4497	25/+	5'-CCTGTGGGAAGTGAAGAAACAACGG-3'
61	cD2-4557	30/-	5'-GCTCCATCTTCCAGTTCAGCCTTTCCCATG-3'
62	cD2-4615	25/-	5'-CTCCGGCTCCAATCTGAGAGTATCC-3' (G, G, A)
63	D2-4746	25/+	5'-CCTAATATCATATGGAGGAGGCTGG-3'
64	D2-4792	25/+	5'-GAAGGAGAAGAAGTCCAGGTATTGG-3'
65	cD2-4922	25/-	5'-CTGTCGACAATTGGAGATCCTGACG-3' (T, T)
66	D2-4994	25/+	5'-GTGGAGCATATGTGAGTGCTATAGC-3'
67	D2-5124	25/+	5'-TCTGACTATGGCCGGAAGGTATCTC-3'
68	D2-5173	25/+	5'-ACATTAATCTTGGCCCCCACTAGAG-3'
69	cD2-5272	19/-	5'-CGATCTCCCGCCCGGTGTG-3' (A)
70	cD2-5318	25/-	5'-CTAACTGGTGATAGCAGCCTCATGG-3'
71	cD2-5656	27/-	5'-CCTACTGAGTTGTATCACTTTCTTTCC-3'
72	cD2-5891	26/-	5'-TGGATTTCTTCCTATTCTCCCTCTTC-3'
73	D2-5770	25/+	5'-TTCAAGGCTGAGAGGGTTATAGACC-3'
74	D2-6152	25/+	5'-TCTGGTTGGCCTACAGAGTGGCAGC-3'
75	cD2-6252	27/-	5'-CCTTCTTTTGTCCAGATTTCCACTTCC-3' (A)

SEQ. ID NO:	PRIMER	MER/SENSE	SEQUENCE
76	D2-6493	35/+	5'-GCGTACAACCATGCTCTCAGTGAACTGCCGGAGAC-3'
77	cD2-6605	24/-	5'-TTCCCAGGGTCATCTTCCCTATAC-3' (G)
78	cD2-6624	31/-	5'-GATGCTAGCCGTGATTATGCAGCACATTCCC-3'
79	D2-6748	25/+	5'-AAACAGAGAACACCCCAAGACAACC-3'
80	cD2-6932	21/-	5'-CGGCATACAGCGTCCATGCTG-3'
81	D2-7055	25/+	5'-GTCTCGGGAAAGGATGGCCATTGTC-3'
82	cD2-7195	25/-	5'-CTCTGGTTGCTTTTGCTTGAAGTCC-3' (A G G)
83	cD2-7217	27/-	5'-CCGCCGCTGCTCTTTTCTGAGCTTCTC-3'
84	D2-7378	25/+	5'-AGGACTACATGGGCTCTGTGTGAGG-3'
85	cD2-7515	19/-	5'-GAGAAGTCCAGCTCCGGCC-3'
86	D2-7769	25/+	5'-AGAGAAACATGGTCACACCAGAAGG-3'
87	cD2-7885	22/-	5'-GTTCTTCGTGTCCTGGTCCTCC-3'
88	D2-8165	25/+	5'-GGAAATATGGAGGAGCCTAGTGAGG-3'
89	cD2-8210	22/-	5'-ACCCAGTACATCTCATGTGTGG-3'
90	D2-8428	28/+	5'-GAGCATGAAACATCATGGCACTATGACC-3'
91	D2-8440	25/+	5'-TCATGGCACTATGACCAAGACCACC-3'
92	cD2-8529	22/-	5'-CAGTCTGACCACTCCGTTCACC-3' (C A G)
93	D2-8773	25/+	5'-AAGGTGAGAAGCAATGCAGCCTTGG-3'
94	D2-8798	29/+	5'-GGGCCATATTCACTGATGAGAACAAGTGG-3'
95	cD2-8865	22/-	5'-TCTTTCCCTGTCAACCAGCTCC-3' (C T)
96	D2-9046	25/+	5'-AATGAAGATCACTGGTTCTCCAGAG-3'
97	D2-9131	25/+	5'-ACGTGAGCAAGAAAGAGGGAGGAGC-3'
98	cD2-9166	22/-	5'-TGTCCCATCCTGCTGTGTCATC-3' (A G)
99	cD2-9234	30/-	5'-GCTAGTTTCTTGTGTTCTCCTTCCATGTGG-3'
100	D2-9344	25/+	5'-TCATATCGAGAAGAGACCAAAGAGG-3'
101	cD2-9429	24/-	5'-ACTCCTTCTCCCTCCATCTGTCTG-3'

SEQ. ID NO:	PRIMER	MER/SENSE	SEQUENCE
102	cD2-9438	27/-	5'-ATGCTTTTGAAGATTCCTTCTCCCTCC-3' A C
103	cD2-9468	32/-	5'-GCACAGCGATTTCTTCTGTGATTGTTAGGTGC-3'
104	D2-9645	25/+	5'-ACAATGGGAACCTTCAAGAGGATGG-3'
105	D2-9656.BAM	45/+	5'-TTATCACATT/GGATCC/TTCAAGAGGATGGA ATGATTGGACACAAG-3' (5'-Fill/BamHI/DEN-2 Sequence)
106	cD2-9668	28/-	5'-CAGAAGGGCACTTGTGTCCAATCATTCC-3'
107	cD2-9779	21/-	5'-CTCCCTGGGAAATTCGGGCTC-3' T G
108	cD2-9796	28/-	5'-CCGTCTCCCGCAAAGACCACCCTGCTCC-3'
109	cD2-9796.XBA	44/-	5'-TTATCACCTA/TCTAGA/CCGTCTCCC GCAAAGACCACCCTGCTCC-3'
110	cD2-9913	26/-	5'-GTTGGAACCCAATGTGATGGTACTGC-3'
111	D2-9937	25/+	5'-ACAAGTCGAACAACCTGGTCCATAC-3'
112	cD2-9977	21/-	5'-GCATGTCTTCCGTCGTCATCC-3' T
113	cD2-10003	25/-	5'-CTTGAATCCACACCCTGTTCCAGAC-3'
114	D2-10203	25/+	5'-ATACACAGATTACATGCCATCCATG-3'
115	cD2-10261	21/-	5'-TTTTGCCTTCTACCACAGGAC-3' T A
116	D2-10289	25/-	5'-GAAACAAGGCTAGAAGTCAGGTCGG-3'
117	cD2-10337	23/-	5'-GACGGGGCTCACAGGTAGCATAG-3'
118	D2-10418	25/+	5'-GCCTGTAGCTCCACCTGAGAAGGTG-3'
119	D2-10470	25/+	5'-GGAAGCTGTACGCATGGCGTAGTGG-3'
120	cD2-10530	19/-	5'-GGGCCCCCGTTGTTGCTGC-3' A
121	cD2-10687	59/-	5'-AGAACCTGTTGATTCAACAGCACCATTCCATTTTCTG-3'
122	cD2-10687.XBA	59/-	5'-TTATCACCTA/GCATGC/TCTAGA/ AGAACCTGTTGATTCAACAGCACCATTCCATTTTCTG-3' (5'-Fill/SphI/XbaI/ 3'-End DEN-2 Sequence)
123	cD2-10687.X2	52/-	5'-TTATCACCTA/TCTAGA/ GAACCTGTTGATTCAACAGCACCATTCCATTTTCTG-3' (5'-Fill/XbaI/ 3'-End DEN-2 Sequence)

While particular embodiments of the invention have been described in detail, it will be apparent to those skilled in the art that these embodiments are exemplary rather than limiting, and the true scope of the invention is that defined within the attached claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: MAHIDOL UNIVERSITY
Bangkok, Thailand

The United States of
America, as represented by the Secretary,
Department of Health and Human Services
c/o Centers for Disease Control and
Prevention
Technology Transfer Office
Mail Stop E-67
1600 Clifton Road
Atlanta, Georgia 30333

(ii) TITLE OF THE INVENTION: INFECTIOUS CDNA CLONES FOR DENGUE 2 VIRUS ...

(iii) NUMBER OF SEQUENCES: 137

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: NEEDLE & ROSENBERG, P.C.
(B) STREET: Suite 1200, 127 Peachtree Street, NE
(C) CITY: Atlanta
(D) STATE: GA
(E) COUNTRY: USA
(F) ZIP: 30303

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: U.S. Serial No. 08/483,292
(B) FILING DATE: 7 Jun 1995
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Spratt, Gwendolyn D.
(B) REGISTRATION NUMBER: 36,016
(C) REFERENCE/DOCKET NUMBER: 14114.0179/P

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 404-688-0770
(B) TELEFAX: 404-688-9880
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10723 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 97...10269
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
AGTTGTTAGT CTACGTGGAC CGACAAAGAC AGATTCTTTG AGGGAGCTAA GCTCAACGTA     60
GTTCTAACAG TTTTTTAATT AGAGAGCAGA TCTCTG ATG AAT AAC CAA CGG AAA      114
                                        Met Asn Asn Gln Arg Lys
                                         1               5

AAG GCG AAA AAC ACG CCT TTC AAT ATG CTG AAA CGC GAG AGA AAC CGC      162
Lys Ala Lys Asn Thr Pro Phe Asn Met Leu Lys Arg Glu Arg Asn Arg
            10                  15                  20

GTG TCG ACT GTG CAA CAG CTG ACA AAG AGA TTC TCA CTT GGA ATG CTG      210
Val Ser Thr Val Gln Gln Leu Thr Lys Arg Phe Ser Leu Gly Met Leu
        25                  30                  35

CAG GGA CGA GGA CCA TTA AAA CTG TTC ATG GCC CTG GTG GCG TTC CTT      258
Gln Gly Arg Gly Pro Leu Lys Leu Phe Met Ala Leu Val Ala Phe Leu
    40                  45                  50

CGT TTC CTA ACA ATC CCA CCA ACA GCA GGG ATA TTG AAG AGA TGG GGA      306
Arg Phe Leu Thr Ile Pro Pro Thr Ala Gly Ile Leu Lys Arg Trp Gly
55                  60                  65                  70

ACA ATT AAA AAA TCA AAA GCT ATT AAT GTT TTG AGA GGG TTC AGG AAA      354
Thr Ile Lys Lys Ser Lys Ala Ile Asn Val Leu Arg Gly Phe Arg Lys
                75                  80                  85

GAG ATT GGA AGG ATG CTG AAC ATC TTG AAT AGG AGA CGC AGA TCT GCA      402
Glu Ile Gly Arg Met Leu Asn Ile Leu Asn Arg Arg Arg Arg Ser Ala
            90                  95                  100

GGC ATG ATC ATT ATG CTG ATT CCA ACA GTG ATG GCG TTC CAT TTA ACC      450
Gly Met Ile Ile Met Leu Ile Pro Thr Val Met Ala Phe His Leu Thr
        105                 110                 115

ACA CGT AAC GGA GAA CCA CAC ATG ATC GTC AGC AGA CAA GAG AAA GGG      498
Thr Arg Asn Gly Glu Pro His Met Ile Val Ser Arg Gln Glu Lys Gly
    120                 125                 130
```

```
AAA AGT CTT CTG TTT AAA ACA GAG GAT GGC GTG AAC ATG TGT ACC CTC      546
Lys Ser Leu Leu Phe Lys Thr Glu Asp Gly Val Asn Met Cys Thr Leu
135                 140                 145                 150

ATG GCC ATG GAC CTT GGT GAA TTG TGT GAA GAC ACA ATC ACG TAC AAG      594
Met Ala Met Asp Leu Gly Glu Leu Cys Glu Asp Thr Ile Thr Tyr Lys
                155                 160                 165

TGT CCC CTT CTC AGG CAG AAT GAG CCA GAA GAC ATA GAC TGT TGG TGC      642
Cys Pro Leu Leu Arg Gln Asn Glu Pro Glu Asp Ile Asp Cys Trp Cys
            170                 175                 180

AAC TCT ACG TCC ACG TGG GTA ACT TAT GGG ACG TGT ACC ACC ATG GGA      690
Asn Ser Thr Ser Thr Trp Val Thr Tyr Gly Thr Cys Thr Thr Met Gly
        185                 190                 195

GAA CAT AGA AGA GAA AAA AGA TCA GTG GCA CTC GTT CCA CAT GTG GGA      738
Glu His Arg Arg Glu Lys Arg Ser Val Ala Leu Val Pro His Val Gly
    200                 205                 210

ATG GGA CTG GAG ACA CGA ACT GAA ACA TGG ATG TCA TCA GAA GGG GCC      786
Met Gly Leu Glu Thr Arg Thr Glu Thr Trp Met Ser Ser Glu Gly Ala
215                 220                 225                 230

TGG AAA CAT GTC CAG AGA ATT GAA ACT TGG ATC TTG AGA CAT CCA GGC      834
Trp Lys His Val Gln Arg Ile Glu Thr Trp Ile Leu Arg His Pro Gly
                235                 240                 245

TTC ACC ATG ATG GCA GCA ATC CTG GCA TAC ACC ATA GGA ACG ACA CAT      882
Phe Thr Met Met Ala Ala Ile Leu Ala Tyr Thr Ile Gly Thr Thr His
            250                 255                 260

TTC CAA AGA GCC CTG ATT TTC ATC TTA CTG ACA GCT GTC ACT CCT TCA      930
Phe Gln Arg Ala Leu Ile Phe Ile Leu Leu Thr Ala Val Thr Pro Ser
        265                 270                 275

ATG ACA ATG CGT TGC ATA GGA ATG TCA AAT AGA GAC TTT GTG GAA GGG      978
Met Thr Met Arg Cys Ile Gly Met Ser Asn Arg Asp Phe Val Glu Gly
    280                 285                 290

GTT TCA GGA GGA AGC TGG GTT GAC ATA GTC TTA GAA CAT GGA AGC TGT     1026
Val Ser Gly Gly Ser Trp Val Asp Ile Val Leu Glu His Gly Ser Cys
295                 300                 305                 310

GTG ACG ACG ATG GCA AAA AAC AAA CCA ACA TTG GAT TTT GAA CTG ATA     1074
Val Thr Thr Met Ala Lys Asn Lys Pro Thr Leu Asp Phe Glu Leu Ile
                315                 320                 325

AAA ACA GAA GCC AAA CAG CCT GCC ACC CTA AGG AAG TAC TGT ATA GAG     1122
Lys Thr Glu Ala Lys Gln Pro Ala Thr Leu Arg Lys Tyr Cys Ile Glu
            330                 335                 340

GCA AAG CTA ACC AAC ACA ACA ACA GAA TCT CGC TGC CCA ACA CAA GGG     1170
Ala Lys Leu Thr Asn Thr Thr Thr Glu Ser Arg Cys Pro Thr Gln Gly
        345                 350                 355
```

```
GAA CCC AGC CTA AAT GAA GAG CAG GAC AAA AGG TTC GTC TGC AAA CAC     1218
Glu Pro Ser Leu Asn Glu Glu Gln Asp Lys Arg Phe Val Cys Lys His
    360                 365                 370

TCC ATG GTA GAC AGA GGA TGG GGA AAT GGA TGT GGA CTA TTT GGA AAG     1266
Ser Met Val Asp Arg Gly Trp Gly Asn Gly Cys Gly Leu Phe Gly Lys
375                 380                 385                 390

GGA GGC ATT GTG ACC TGT GCT ATG TTC AGA TGC AAA AAG AAC ATG GAA     1314
Gly Gly Ile Val Thr Cys Ala Met Phe Arg Cys Lys Lys Asn Met Glu
                395                 400                 405

GGA AAA GTT GTG CAA CCA GAA AAC TTG GAA TAC ACC ATT GTG ATA ACA     1362
Gly Lys Val Val Gln Pro Glu Asn Leu Glu Tyr Thr Ile Val Ile Thr
            410                 415                 420

CCT CAC TCA GGG GAA GAG CAT GCA GTC GGA AAT GAC ACA GGA AAA CAT     1410
Pro His Ser Gly Glu Glu His Ala Val Gly Asn Asp Thr Gly Lys His
        425                 430                 435

GGC AAG GAA ATC AAA ATA ACA CCA CAG AGT TCC ATC ACA GAA GCA GAA     1458
Gly Lys Glu Ile Lys Ile Thr Pro Gln Ser Ser Ile Thr Glu Ala Glu
    440                 445                 450

TTG ACA GGT TAT GGC ACT GTC ACA ATG GAG TGC TCT CCA AGA ACG GGC     1506
Leu Thr Gly Tyr Gly Thr Val Thr Met Glu Cys Ser Pro Arg Thr Gly
455                 460                 465                 470

CTC GAC TTC AAT GAG ATG GTG TTG CTG CAG ATG GAA AAT AAA GCT TGG     1554
Leu Asp Phe Asn Glu Met Val Leu Leu Gln Met Glu Asn Lys Ala Trp
                475                 480                 485

CTG GTG CAC AGG CAA TGG TTC CTA GAC CTG CCG TTA CCA TGG TTG CCC     1602
Leu Val His Arg Gln Trp Phe Leu Asp Leu Pro Leu Pro Trp Leu Pro
            490                 495                 500

GGA GCG GAC ACA CAA GGG TCA AAT TGG ATA CAG AAA GAG ACA TTG GTC     1650
Gly Ala Asp Thr Gln Gly Ser Asn Trp Ile Gln Lys Glu Thr Leu Val
        505                 510                 515

ACT TTC AAA AAT CCC CAT GCG AAG AAA CAG GAT GTT GTT GTT TTA GGA     1698
Thr Phe Lys Asn Pro His Ala Lys Lys Gln Asp Val Val Val Leu Gly
    520                 525                 530

TCC CAA GAA GGG GCC ATG CAC ACA GCA CTT ACA GGG GCC ACA GAA ATC     1746
Ser Gln Glu Gly Ala Met His Thr Ala Leu Thr Gly Ala Thr Glu Ile
535                 540                 545                 550

CAA ATG TCA TCA GGA AAC TTA CTC TTC ACA GGA CAT CTC AAG TGC AGG     1794
Gln Met Ser Ser Gly Asn Leu Leu Phe Thr Gly His Leu Lys Cys Arg
                555                 560                 565

CTG AGA ATG GAC AAG CTA CAG CTC AAA GGA ATG TCA TAC TCT ATG TGC     1842
Leu Arg Met Asp Lys Leu Gln Leu Lys Gly Met Ser Tyr Ser Met Cys
            570                 575                 580

ACA GGA AAG TTT AAA GTT GTG AAG GAA ATA GCA GAA ACA CAA CAT GGA     1890
Thr Gly Lys Phe Lys Val Val Lys Glu Ile Ala Glu Thr Gln His Gly
        585                 590                 595
```

```
ACA ATA GTT ATC AGA GTG CAA TAT GAA GGG GAC GGC TCT CCA TGC AAG    1938
Thr Ile Val Ile Arg Val Gln Tyr Glu Gly Asp Gly Ser Pro Cys Lys
    600                 605                 610

ATC CCT TTT GAG ATA ATG GAT TTG GAA AAA AGA CAT GTC TTA GGT CGC    1986
Ile Pro Phe Glu Ile Met Asp Leu Glu Lys Arg His Val Leu Gly Arg
615                 620                 625                 630

CTG ATT ACA GTC AAC CCA ATT GTG ACA GAA AAA GAT AGC CCA GTC AAC    2034
Leu Ile Thr Val Asn Pro Ile Val Thr Glu Lys Asp Ser Pro Val Asn
                635                 640                 645

ATA GAA GCA GAA CCT CCA TTC GGA GAC AGC TAC ATC ATC ATA GGA GTA    2082
Ile Glu Ala Glu Pro Pro Phe Gly Asp Ser Tyr Ile Ile Ile Gly Val
            650                 655                 660

GAG CCG GGA CAA CTG AAG CTC AAC TGG TTT AAG AAA GGA AGT TCT ATC    2130
Glu Pro Gly Gln Leu Lys Leu Asn Trp Phe Lys Lys Gly Ser Ser Ile
        665                 670                 675

GGC CAA ATG TTT GAG ACA ACA ATG AGG GGG GCG AAG AGA ATG GCC ATT    2178
Gly Gln Met Phe Glu Thr Thr Met Arg Gly Ala Lys Arg Met Ala Ile
    680                 685                 690

TTA GGT GAC ACA GCC TGG GAT TTT GGA TCC TTG GGA GGA GTG TTT ACA    2226
Leu Gly Asp Thr Ala Trp Asp Phe Gly Ser Leu Gly Gly Val Phe Thr
695                 700                 705                 710

TCT ATA GGA AAG GCT CTC CAC CAA GTC TTT GGA GCA ATC TAT GGA GCT    2274
Ser Ile Gly Lys Ala Leu His Gln Val Phe Gly Ala Ile Tyr Gly Ala
                715                 720                 725

GCC TTC AGT GGG GTT TCA TGG ACT ATG AAA ATC CTC ATA GGA GTC ATT    2322
Ala Phe Ser Gly Val Ser Trp Thr Met Lys Ile Leu Ile Gly Val Ile
            730                 735                 740

ATC ACA TGG ATA GGA ATG AAT TCA CGC AGC ACC TCA CTG TCT GTG ACA    2370
Ile Thr Trp Ile Gly Met Asn Ser Arg Ser Thr Ser Leu Ser Val Thr
        745                 750                 755

CTA GTA TTG GTG GGA ATT GTG ACA CTG TAT TTG GGA GTC ATG GTG CAG    2418
Leu Val Leu Val Gly Ile Val Thr Leu Tyr Leu Gly Val Met Val Gln
    760                 765                 770

GCC GAT AGT GGT TGC GTT GTG AGC TGG AAA AAC AAA GAA CTG AAA TGT    2466
Ala Asp Ser Gly Cys Val Val Ser Trp Lys Asn Lys Glu Leu Lys Cys
775                 780                 785                 790

GGC AGT GGG ATT TTC ATC ACA GAC AAC GTG CAC ACA TGG ACA GAA CAA    2514
Gly Ser Gly Ile Phe Ile Thr Asp Asn Val His Thr Trp Thr Glu Gln
                795                 800                 805

TAC AAG TTC CAA CCA GAA TCC CCT TCA AAA CTA GCT TCA GCT ATC CAG    2562
Tyr Lys Phe Gln Pro Glu Ser Pro Ser Lys Leu Ala Ser Ala Ile Gln
            810                 815                 820

AAA GCC CAT GAA GAG GGC ATT TGT GGA ATC CGC TCA GTA ACA AGA CTG    2610
Lys Ala His Glu Glu Gly Ile Cys Gly Ile Arg Ser Val Thr Arg Leu
        825                 830                 835
```

```
GAG AAT CTG ATG TGG AAA CAA ATA ACA CCA GAA TTG AAT CAC ATT CTA     2658
Glu Asn Leu Met Trp Lys Gln Ile Thr Pro Glu Leu Asn His Ile Leu
    840                 845                 850

TCA GAA AAT GAG GTG AAG TTA ACT ATT ATG ACA GGA GAC ATC AAA GGA     2706
Ser Glu Asn Glu Val Lys Leu Thr Ile Met Thr Gly Asp Ile Lys Gly
855                 860                 865                 870

ATC ATG CAG GCA GGA AAA CGA TCT CTG CGG CCT CAG CCC ACT GAG CTG     2754
Ile Met Gln Ala Gly Lys Arg Ser Leu Arg Pro Gln Pro Thr Glu Leu
                875                 880                 885

AAG TAT TCA TGG AAA ACA TGG GGC AAA GCA AAA ATG CTC TCT ACA GAG     2802
Lys Tyr Ser Trp Lys Thr Trp Gly Lys Ala Lys Met Leu Ser Thr Glu
            890                 895                 900

TCT CAT AAC CAG ACC TTT CTC ATT GAT GGC CCC GAA ACA GCA GAA TGC     2850
Ser His Asn Gln Thr Phe Leu Ile Asp Gly Pro Glu Thr Ala Glu Cys
        905                 910                 915

CCC AAC ACA AAT AGA GCT TGG AAT TCG TTG GAA GTT GAA GAC TAT GGC     2898
Pro Asn Thr Asn Arg Ala Trp Asn Ser Leu Glu Val Glu Asp Tyr Gly
    920                 925                 930

TTT GGA GTA TTC ACC ACC AAT ATA TGG CTA AAA TTG AAA GAA AAA CAG     2946
Phe Gly Val Phe Thr Thr Asn Ile Trp Leu Lys Leu Lys Glu Lys Gln
935                 940                 945                 950

GAT GTA TTC TGC GAC TCA AAA CTC ATG TCA GCG GCC ATA AAA GAC AAC     2994
Asp Val Phe Cys Asp Ser Lys Leu Met Ser Ala Ala Ile Lys Asp Asn
                955                 960                 965

AGA GCC GTC CAT GCC GAT ATG GGT TAT TGG ATA GAA AGT GCA CTC AAT     3042
Arg Ala Val His Ala Asp Met Gly Tyr Trp Ile Glu Ser Ala Leu Asn
            970                 975                 980

GAC ACA TGG AAG ATA GAG AAA GCC TCT TTC ATT GAA GTT AAA AAC TGC     3090
Asp Thr Trp Lys Ile Glu Lys Ala Ser Phe Ile Glu Val Lys Asn Cys
        985                 990                 995

CAC TGG CCA AAA TCA CAC ACC CTC TGG AGC AAT GGA GTG CTA GAA AGT     3138
His Trp Pro Lys Ser His Thr Leu Trp Ser Asn Gly Val Leu Glu Ser
   1000                1005                1010

GAG ATG ATA ATT CCA AAG AAT CTC GCT GGA CCA GTG TCT CAA CAC AAC     3186
Glu Met Ile Ile Pro Lys Asn Leu Ala Gly Pro Val Ser Gln His Asn
1015                1020                1025                1030

TAT AGA CCA GGC TAC CAT ACA CAA ATA ACA GGA CCA TGG CAT CTA GGT     3234
Tyr Arg Pro Gly Tyr His Thr Gln Ile Thr Gly Pro Trp His Leu Gly
                1035                1040                1045

AAG CTT GAG ATG GAC TTT GAT TTC TGT GAT GGA ACA ACA GTG GTA GTG     3282
Lys Leu Glu Met Asp Phe Asp Phe Cys Asp Gly Thr Thr Val Val Val
            1050                1055                1060

ACT GAG GAC TGC GGA AAT AGA GGA CCC TCT TTG AGA ACA ACC ACT GCC     3330
Thr Glu Asp Cys Gly Asn Arg Gly Pro Ser Leu Arg Thr Thr Thr Ala
        1065                1070                1075
```

```
TCT GGA AAA CTC ATA ACA GAA TGG TGC TGC CGA TCT TGC ACA TTA CCA     3378
Ser Gly Lys Leu Ile Thr Glu Trp Cys Cys Arg Ser Cys Thr Leu Pro
   1080                1085                1090

CCG CTA AGA TAC AGA GGT GAG GAT GGG TGC TGG TAC GGG ATG GAA ATC     3426
Pro Leu Arg Tyr Arg Gly Glu Asp Gly Cys Trp Tyr Gly Met Glu Ile
1095                1100                1105                1110

AGA CCA TTG AAG GAG AAA GAA GAG AAT TTG GTC AAC TCC TTG GTC ACA     3474
Arg Pro Leu Lys Glu Lys Glu Glu Asn Leu Val Asn Ser Leu Val Thr
                1115                1120                1125

GCT GGA CAT GGG CAG GTC GAC AAC TTT TCA CTA GGA GTC TTG GGA ATG     3522
Ala Gly His Gly Gln Val Asp Asn Phe Ser Leu Gly Val Leu Gly Met
            1130                1135                1140

GCA TTG TTC CTG GAG GAA ATG CTT AGG ACC CGA GTA GGA ACG AAA CAT     3570
Ala Leu Phe Leu Glu Glu Met Leu Arg Thr Arg Val Gly Thr Lys His
        1145                1150                1155

GCA ATA CTA CTA GTT GCA GTT TCT TTT GTG ACA TTG ATC ACA GGG AAC     3618
Ala Ile Leu Leu Val Ala Val Ser Phe Val Thr Leu Ile Thr Gly Asn
   1160                1165                1170

ATG TCC TTT AGA GAC CTG GGA AGA GTG ATG GTT ATG GTA GGC GCC ACT     3666
Met Ser Phe Arg Asp Leu Gly Arg Val Met Val Met Val Gly Ala Thr
1175                1180                1185                1190

ATG ACG GAT GAC ATA GGT ATG GGC GTG ACT TAT CTT GCC CTA CTA GCA     3714
Met Thr Asp Asp Ile Gly Met Gly Val Thr Tyr Leu Ala Leu Leu Ala
                1195                1200                1205

GCC TTC AAA GTC AGA CCA ACT TTT GCA GCT GGA CTA CTC TTG AGA AAG     3762
Ala Phe Lys Val Arg Pro Thr Phe Ala Ala Gly Leu Leu Leu Arg Lys
            1210                1215                1220

CTG ACC TCC AAG GAA TTG ATG ATG ACT ACT ATA GGA ATT GTA CTC CTC     3810
Leu Thr Ser Lys Glu Leu Met Met Thr Thr Ile Gly Ile Val Leu Leu
        1225                1230                1235

TCC CAG AGC ACC ATA CCA GAG ACC ATT CTT GAG TTG ACT GAT GCG TTA     3858
Ser Gln Ser Thr Ile Pro Glu Thr Ile Leu Glu Leu Thr Asp Ala Leu
   1240                1245                1250

GCC TTA GGC ATG ATG GTC CTC AAA ATG GTG AGA AAT ATG GAA AAG TAT     3906
Ala Leu Gly Met Met Val Leu Lys Met Val Arg Asn Met Glu Lys Tyr
1255                1260                1265                1270

CAA TTG GCA GTG ACT ATC ATG GCT ATC TTG TGC GTC CCA AAC GCA GTG     3954
Gln Leu Ala Val Thr Ile Met Ala Ile Leu Cys Val Pro Asn Ala Val
                1275                1280                1285

ATA TTA CAA AAC GCA TGG AAA GTG AGT TGC ACA ATA TTG GCA GTG GTG     4002
Ile Leu Gln Asn Ala Trp Lys Val Ser Cys Thr Ile Leu Ala Val Val
            1290                1295                1300

TCC GTT TCC CCA CTG CTC TTA ACA TCC TCA CAG CAA AAA ACA GAT TGG     4050
Ser Val Ser Pro Leu Leu Leu Thr Ser Ser Gln Gln Lys Thr Asp Trp
        1305                1310                1315
```

```
ATA CCA TTA GCA TTG ACG ATC AAA GGT CTC AAT CCA ACA GCT ATT TTT    4098
Ile Pro Leu Ala Leu Thr Ile Lys Gly Leu Asn Pro Thr Ala Ile Phe
   1320                1325                1330

CTA ACA ACC CTC TCA AGA ACC AGC AAG AAA AGG AGC TGG CCA TTA AAT    4146
Leu Thr Thr Leu Ser Arg Thr Ser Lys Lys Arg Ser Trp Pro Leu Asn
1335                1340                1345                1350

GAG GCT ATC ATG GCA GTC GGG ATG GTG AGC ATT TTA GCC AGT TCT CTC    4194
Glu Ala Ile Met Ala Val Gly Met Val Ser Ile Leu Ala Ser Ser Leu
                1355                1360                1365

CTA AAA AAT GAT ATT CCC ATG ACA GGA CCA TTA GTG GCT GGA GGG CTC    4242
Leu Lys Asn Asp Ile Pro Met Thr Gly Pro Leu Val Ala Gly Gly Leu
            1370                1375                1380

CTC ACT GTG TGC TAC GTG CTC ACT GGA CGA TCG GCC GAT TTG GAA CTG    4290
Leu Thr Val Cys Tyr Val Leu Thr Gly Arg Ser Ala Asp Leu Glu Leu
        1385                1390                1395

GAG AGA GCA GCC GAT GTC AAA TGG GAA GAC CAG GCA GAG ATA TCA GGA    4338
Glu Arg Ala Ala Asp Val Lys Trp Glu Asp Gln Ala Glu Ile Ser Gly
   1400                1405                1410

AGC AGT CCA ATC CTG TCA ATA ACA ATA TCA GAA GAT GGT AGC ATG TCG    4386
Ser Ser Pro Ile Leu Ser Ile Thr Ile Ser Glu Asp Gly Ser Met Ser
1415                1420                1425                1430

ATA AAA AAT GAA GAG GAA GAA CAA ACA CTG ACC ATA CTC ATT AGA ACA    4434
Ile Lys Asn Glu Glu Glu Glu Gln Thr Leu Thr Ile Leu Ile Arg Thr
                1435                1440                1445

GGA TTG CTG GTG ATC TCA GGA CTT TTT CCT GTA TCA ATA CCA ATC ACG    4482
Gly Leu Leu Val Ile Ser Gly Leu Phe Pro Val Ser Ile Pro Ile Thr
            1450                1455                1460

GCA GCA GCA TGG TAC CTG TGG GAA GTG AAG AAA CAA CGG GCC GGA GTA    4530
Ala Ala Ala Trp Tyr Leu Trp Glu Val Lys Lys Gln Arg Ala Gly Val
        1465                1470                1475

TTG TGG GAT GTT CCT TCA CCC CCA CCC ATG GGA AAG GCT GAA CTG GAA    4578
Leu Trp Asp Val Pro Ser Pro Pro Pro Met Gly Lys Ala Glu Leu Glu
   1480                1485                1490

GAT GGA GCC TAT AGA ATT AAG CAA AAA GGG ATT CTT GGA TAT TCC CAG    4626
Asp Gly Ala Tyr Arg Ile Lys Gln Lys Gly Ile Leu Gly Tyr Ser Gln
1495                1500                1505                1510

ATC GGA GCC GGA GTT TAC AAA GAA GGA ACA TTC CAT ACA ATG TGG CAT    4674
Ile Gly Ala Gly Val Tyr Lys Glu Gly Thr Phe His Thr Met Trp His
                1515                1520                1525

GTC ACA CGT GGC GCT GTT CTA ATG CAT AAA GGA AAG AGG ATT GAA CCA    4722
Val Thr Arg Gly Ala Val Leu Met His Lys Gly Lys Arg Ile Glu Pro
            1530                1535                1540

TCA TGG GCG GAC GTC AAG AAA GAC CTA ATA TCA TAT GGA GGA GGC TGG    4770
Ser Trp Ala Asp Val Lys Lys Asp Leu Ile Ser Tyr Gly Gly Gly Trp
        1545                1550                1555
```

```
AAG TTA GAA GGA GAA TGG AAG GAA GGA GAA GAA GTC CAG GTA TTG GCA    4818
Lys Leu Glu Gly Glu Trp Lys Glu Gly Glu Glu Val Gln Val Leu Ala
   1560                1565                1570

CTG GAG CCT GGA AAA AAT CCA AGA GCC GTC CAA ACG AAA CCT GGT CTT    4866
Leu Glu Pro Gly Lys Asn Pro Arg Ala Val Gln Thr Lys Pro Gly Leu
1575                1580                1585                1590

TTC AAA ACC AAC GCC GGA ACA ATA GGT GCT GTA TCT CTG GAC TTT TCT    4914
Phe Lys Thr Asn Ala Gly Thr Ile Gly Ala Val Ser Leu Asp Phe Ser
                1595                1600                1605

CCT GGA ACG TCA GGA TCT CCA ATT ATC GAC AAA AAA GGA AAA GTT GTG    4962
Pro Gly Thr Ser Gly Ser Pro Ile Ile Asp Lys Lys Gly Lys Val Val
           1610                1615                1620

GGT CTT TAT GGT AAT GGT GTT GTT ACA AGG AGT GGA GCA TAT GTG AGT    5010
Gly Leu Tyr Gly Asn Gly Val Val Thr Arg Ser Gly Ala Tyr Val Ser
      1625                1630                1635

GCT ATA GCC CAG ACT GAA AAA AGC ATT GAA GAC AAC CCA GAG ATC GAA    5058
Ala Ile Ala Gln Thr Glu Lys Ser Ile Glu Asp Asn Pro Glu Ile Glu
   1640                1645                1650

GAT GAC ATT TTC CGA AAG AGA AGA CTG ACC ATC ATG GAC CTC CAC CCA    5106
Asp Asp Ile Phe Arg Lys Arg Arg Leu Thr Ile Met Asp Leu His Pro
1655                1660                1665                1670

GGA GCG GGA AAG ACG AAG AGA TAC CTT CCG GCC ATA GTC AGA GAA GCT    5154
Gly Ala Gly Lys Thr Lys Arg Tyr Leu Pro Ala Ile Val Arg Glu Ala
                1675                1680                1685

ATA AAA CGG GGT TTG AGA ACA TTA ATC TTG GCC CCC ACT AGA GTT GTG    5202
Ile Lys Arg Gly Leu Arg Thr Leu Ile Leu Ala Pro Thr Arg Val Val
           1690                1695                1700

GCA GCT GAA ATG GAG GAA GCC CTT AGA GGA CTT CCA ATA AGA TAC CAG    5250
Ala Ala Glu Met Glu Glu Ala Leu Arg Gly Leu Pro Ile Arg Tyr Gln
      1705                1710                1715

ACC CCA GCC ATC AGA GCT GAG CAC ACC GGG CGG GAG ATT GTG GAC CTA    5298
Thr Pro Ala Ile Arg Ala Glu His Thr Gly Arg Glu Ile Val Asp Leu
   1720                1725                1730

ATG TGT CAT GCC ACA TTT ACC ATG AGG CTG CTA TCA CCA GTT AGA GTG    5346
Met Cys His Ala Thr Phe Thr Met Arg Leu Leu Ser Pro Val Arg Val
1735                1740                1745                1750

CCA AAC TAC AAC CTG ATT ATC ATG GAC GAA GCC CAT TTC ACA GAC CCA    5394
Pro Asn Tyr Asn Leu Ile Ile Met Asp Glu Ala His Phe Thr Asp Pro
                1755                1760                1765

GCA AGT ATA GCA GCT AGA GGA TAC ATC TCA ACT CGA GTG GAG ATG GGT    5442
Ala Ser Ile Ala Ala Arg Gly Tyr Ile Ser Thr Arg Val Glu Met Gly
           1770                1775                1780

GAG GCA GCT GGG ATT TTT ATG ACA GCC ACT CCC CCG GGA AGC AGA GAC    5490
Glu Ala Ala Gly Ile Phe Met Thr Ala Thr Pro Pro Gly Ser Arg Asp
      1785                1790                1795
```

```
CCA TTT CCT CAG AGC AAT GCA CCA ATC ATA GAT GAA GAA AGA GAA ATC    5538
Pro Phe Pro Gln Ser Asn Ala Pro Ile Ile Asp Glu Glu Arg Glu Ile
   1800                1805                1810

CCT GAA CGC TCG TGG AAT TCC GGA CAT GAA TGG GTC ACG GAT TTT AAA    5586
Pro Glu Arg Ser Trp Asn Ser Gly His Glu Trp Val Thr Asp Phe Lys
1815                1820                1825                1830

GGG AAG ACT GTT TGG TTC GTT CCA AGT ATA AAA GCA GGA AAT GAT ATA    5634
Gly Lys Thr Val Trp Phe Val Pro Ser Ile Lys Ala Gly Asn Asp Ile
                1835                1840                1845

GCA GCT TGC CTG AGG AAA AAT GGA AAG AAA GTG ATA CAA CTC AGT AGG    5682
Ala Ala Cys Leu Arg Lys Asn Gly Lys Lys Val Ile Gln Leu Ser Arg
           1850                1855                1860

AAG ACC TTT GAT TCT GAG TAT GTC AAG ACT AGA ACC AAT GAT TGG GAC    5730
Lys Thr Phe Asp Ser Glu Tyr Val Lys Thr Arg Thr Asn Asp Trp Asp
       1865                1870                1875

TTC GTG GTT ACA ACT GAC ATT TCA GAA ATG GGT GCC AAT TTC AAG GCT    5778
Phe Val Val Thr Thr Asp Ile Ser Glu Met Gly Ala Asn Phe Lys Ala
   1880                1885                1890

GAG AGG GTT ATA GAC CCC AGA CGC TGC ATG AAA CCA GTC ATA CTA ACA    5826
Glu Arg Val Ile Asp Pro Arg Arg Cys Met Lys Pro Val Ile Leu Thr
1895                1900                1905                1910

GAT GGT GAA GAG CGG GTG ATT CTG GCA GGA CCT ATG CCA GTG ACC CAC    5874
Asp Gly Glu Glu Arg Val Ile Leu Ala Gly Pro Met Pro Val Thr His
                1915                1920                1925

TCT AGT GCA GCA CAA AGA AGA GGG AGA ATA GGA AGA AAT CCA AAA AAT    5922
Ser Ser Ala Ala Gln Arg Arg Gly Arg Ile Gly Arg Asn Pro Lys Asn
           1930                1935                1940

GAG AAT GAC CAG TAC ATA TAC ATG GGG GAA CCT CTG GAA AAT GAT GAA    5970
Glu Asn Asp Gln Tyr Ile Tyr Met Gly Glu Pro Leu Glu Asn Asp Glu
       1945                1950                1955

GAC TGT GCA CAC TGG AAA GAA GCT AAA ATG CTC CTA GAT AAC ATC AAC    6018
Asp Cys Ala His Trp Lys Glu Ala Lys Met Leu Leu Asp Asn Ile Asn
   1960                1965                1970

ACG CCA GAA GGA ATC ATT CCT AGC ATG TTC GAA CCA GAG CGT GAA AAG    6066
Thr Pro Glu Gly Ile Ile Pro Ser Met Phe Glu Pro Glu Arg Glu Lys
1975                1980                1985                1990

GTG GAT GCC ATT GAT GGC GAA TAC CGC TTG AGA GGA GAA GCA AGG AAA    6114
Val Asp Ala Ile Asp Gly Glu Tyr Arg Leu Arg Gly Glu Ala Arg Lys
                1995                2000                2005

ACC TTT GTA GAC TTA ATG AGA AGA GGA GAC CTA CCA GTC TGG TTG GCC    6162
Thr Phe Val Asp Leu Met Arg Arg Gly Asp Leu Pro Val Trp Leu Ala
           2010                2015                2020

TAC AGA GTG GCA GCT GAA GGC ATC AAC TAC GCA GAC AGA AGG TGG TGT    6210
Tyr Arg Val Ala Ala Glu Gly Ile Asn Tyr Ala Asp Arg Arg Trp Cys
       2025                2030                2035
```

```
TTT GAT GGA GTC AAG AAC AAC CAA ATC CTA GAA GAA AAC GTG GAA GTT    6258
Phe Asp Gly Val Lys Asn Asn Gln Ile Leu Glu Glu Asn Val Glu Val
   2040                2045                2050

GAA ATC TGG ACA AAA GAA GGG GAA AGG AAG AAA TTG AAA CCC AGA TGG    6306
Glu Ile Trp Thr Lys Glu Gly Glu Arg Lys Lys Leu Lys Pro Arg Trp
2055                2060                2065                2070

TTG GAT GCT AGG ATC TAT TCT GAC CCA CTG GCG CTA AAA GAA TTT AAG    6354
Leu Asp Ala Arg Ile Tyr Ser Asp Pro Leu Ala Leu Lys Glu Phe Lys
                2075                2080                2085

GAA TTT GCA GCC GGA AGA AAG TCT CTG ACC CTG AAC CTA ATC ACA GAA    6402
Glu Phe Ala Ala Gly Arg Lys Ser Leu Thr Leu Asn Leu Ile Thr Glu
           2090                2095                2100

ATG GGT AGG CTC CCA ACC TTC ATG ACT CAG AAG GCA AGA GAC GCA CTG    6450
Met Gly Arg Leu Pro Thr Phe Met Thr Gln Lys Ala Arg Asp Ala Leu
       2105                2110                2115

GAC AAC TTA GCA GTG CTG CAC ACG GCT GAG GCA GGT GGA AGG GCG TAC    6498
Asp Asn Leu Ala Val Leu His Thr Ala Glu Ala Gly Gly Arg Ala Tyr
   2120                2125                2130

AAC CAT GCT CTC AGT GAA CTG CCG GAG ACC CTG GAG ACA TTG CTT TTA    6546
Asn His Ala Leu Ser Glu Leu Pro Glu Thr Leu Glu Thr Leu Leu Leu
2135                2140                2145                2150

CTG ACA CTT CTG GCT ACA GTC ACG GGA GGG ATC TTT TTA TTC TTG ATG    6594
Leu Thr Leu Leu Ala Thr Val Thr Gly Gly Ile Phe Leu Phe Leu Met
                2155                2160                2165

AGC GGA AGG GGC ATA GGG AAG ATG ACC CTG GGA ATG TGC TGC ATA ATC    6642
Ser Gly Arg Gly Ile Gly Lys Met Thr Leu Gly Met Cys Cys Ile Ile
           2170                2175                2180

ACG GCT AGC ATC CTC CTA TGG TAC GCA CAA ATA CAG CCA CAC TGG ATA    6690
Thr Ala Ser Ile Leu Leu Trp Tyr Ala Gln Ile Gln Pro His Trp Ile
       2185                2190                2195

GCA GCT TCA ATA ATA CTG GAG TTT TTT CTC ATA GTT TTG CTT ATT CCA    6738
Ala Ala Ser Ile Ile Leu Glu Phe Phe Leu Ile Val Leu Leu Ile Pro
   2200                2205                2210

GAA CCT GAA AAA CAG AGA ACA CCC CAA GAC AAC CAA CTG ACC TAC GTT    6786
Glu Pro Glu Lys Gln Arg Thr Pro Gln Asp Asn Gln Leu Thr Tyr Val
2215                2220                2225                2230

GTC ATA GCC ATC CTC ACA GTG GTG GCC GCA ACC ATG GCA AAC GAG ATG    6834
Val Ile Ala Ile Leu Thr Val Val Ala Ala Thr Met Ala Asn Glu Met
                2235                2240                2245

GGT TTC CTA GAA AAA ACG AAG AAA GAT CTC GGA TTG GGA AGC ATT GCA    6882
Gly Phe Leu Glu Lys Thr Lys Lys Asp Leu Gly Leu Gly Ser Ile Ala
           2250                2255                2260

ACC CAG CAA CCC GAG AGC AAC ATC CTG GAC ATA GAT CTA CGT CCT GCA    6930
Thr Gln Gln Pro Glu Ser Asn Ile Leu Asp Ile Asp Leu Arg Pro Ala
       2265                2270                2275
```

```
TCA GCA TGG ACG CTG TAT GCC GTG GCC ACA ACA TTT GTT ACA CCA ATG     6978
Ser Ala Trp Thr Leu Tyr Ala Val Ala Thr Thr Phe Val Thr Pro Met
   2280                2285                2290

TTG AGA CAT AGC ATT GAA AAT TCC TCA GTG AAT GTG TCC CTA ACA GCT     7026
Leu Arg His Ser Ile Glu Asn Ser Ser Val Asn Val Ser Leu Thr Ala
2295                2300                2305                2310

ATA GCC AAC CAA GCC ACA GTG TTA ATG GGT CTC GGG AAA GGA TGG CCA     7074
Ile Ala Asn Gln Ala Thr Val Leu Met Gly Leu Gly Lys Gly Trp Pro
                2315                2320                2325

TTG TCA AAG ATG GAC ATC GGA GTT CCC CTT CTC GCC ATT GGA TGC TAC     7122
Leu Ser Lys Met Asp Ile Gly Val Pro Leu Leu Ala Ile Gly Cys Tyr
           2330                2335                2340

TCA CAA GTC AAC CCC ATA ACT CTC ACA GCA GCT CTT TTC TTA TTG GTA     7170
Ser Gln Val Asn Pro Ile Thr Leu Thr Ala Ala Leu Phe Leu Leu Val
      2345                2350                2355

GCA CAT TAT GCC ATC ATA GGG CCA GGA CTC CAA GCA AAA GCA ACC AGA     7218
Ala His Tyr Ala Ile Ile Gly Pro Gly Leu Gln Ala Lys Ala Thr Arg
   2360                2365                2370

GAA GCT CAG AAA AGA GCA GCG GCG GGC ATC ATG AAA AAC CCA ACT GTC     7266
Glu Ala Gln Lys Arg Ala Ala Ala Gly Ile Met Lys Asn Pro Thr Val
2375                2380                2385                2390

GAT GGA ATA ACA GTG ATT GAC CTA GAT CCA ATA CCT TAT GAT CCA AAG     7314
Asp Gly Ile Thr Val Ile Asp Leu Asp Pro Ile Pro Tyr Asp Pro Lys
                2395                2400                2405

TTT GAA AAG CAG TTG GGA CAA GTA ATG CTC CTA GTC CTC TGC GTG ACT     7362
Phe Glu Lys Gln Leu Gly Gln Val Met Leu Leu Val Leu Cys Val Thr
           2410                2415                2420

CAA GTA TTG ATG ATG AGG ACT ACA TGG GCT CTG TGT GAG GCT TTA ACC     7410
Gln Val Leu Met Met Arg Thr Thr Trp Ala Leu Cys Glu Ala Leu Thr
      2425                2430                2435

TTA GCT ACC GGG CCC ATC TCC ACA TTG TGG GAA GGA AAT CCA GGG AGG     7458
Leu Ala Thr Gly Pro Ile Ser Thr Leu Trp Glu Gly Asn Pro Gly Arg
   2440                2445                2450

TTT TGG AAC ACT ACC ATT GCG GTG TCA ATG GCT AAC ATT TTT AGA GGG     7506
Phe Trp Asn Thr Thr Ile Ala Val Ser Met Ala Asn Ile Phe Arg Gly
2455                2460                2465                2470

AGT TAC TTG GCC GGA GCT GGA CTT CTC TTT TCT ATT ATG AAG AAC ACA     7554
Ser Tyr Leu Ala Gly Ala Gly Leu Leu Phe Ser Ile Met Lys Asn Thr
                2475                2480                2485

ACC AAC ACA AGA AGG GGA ACT GGC AAC ATA GGA GAG ACG CTT GGA GAG     7602
Thr Asn Thr Arg Arg Gly Thr Gly Asn Ile Gly Glu Thr Leu Gly Glu
           2490                2495                2500

AAA TGG AAA AGC CGA TTG AAC GCA TTG GGA AAA AGT GAA TTC CAG ATC     7650
Lys Trp Lys Ser Arg Leu Asn Ala Leu Gly Lys Ser Glu Phe Gln Ile
      2505                2510                2515
```

```
TAC AAG AAA AGT GGA ATC CAG GAA GTG GAT AGA ACC TTA GCA AAA GAA     7698
Tyr Lys Lys Ser Gly Ile Gln Glu Val Asp Arg Thr Leu Ala Lys Glu
   2520                2525                2530

GGC ATT AAA AGA GGA GAA ACG GAC CAT CAC GCT GTG TCG CGA GGC TCA     7746
Gly Ile Lys Arg Gly Glu Thr Asp His His Ala Val Ser Arg Gly Ser
2535                2540                2545                2550

GCA AAA CTG AGA TGG TTC GTT GAG AGA AAC ATG GTC ACA CCA GAA GGG     7794
Ala Lys Leu Arg Trp Phe Val Glu Arg Asn Met Val Thr Pro Glu Gly
                2555                2560                2565

AAA GTA GTG GAC CTC GGT TGT GGC AGA GGA GGC TGG TCA TAC TAT TGT     7842
Lys Val Val Asp Leu Gly Cys Gly Arg Gly Gly Trp Ser Tyr Tyr Cys
            2570                2575                2580

GGA GGA CTA AAG AAT GTA AGA GAA GTC AAA GGC CTA ACA AAA GGA GGA     7890
Gly Gly Leu Lys Asn Val Arg Glu Val Lys Gly Leu Thr Lys Gly Gly
        2585                2590                2595

CCA GGA CAC GAA GAA CCC ATC CCC ATG TCA ACA TAT GGG TGG AAT CTA     7938
Pro Gly His Glu Glu Pro Ile Pro Met Ser Thr Tyr Gly Trp Asn Leu
   2600                2605                2610

GTG CGT CTT CAA AGT GGA GTT GAC GTT TTC TTC ATC CCG CCA GAA AAG     7986
Val Arg Leu Gln Ser Gly Val Asp Val Phe Phe Ile Pro Pro Glu Lys
2615                2620                2625                2630

TGT GAC ACA TTA TTG TGT GAC ATA GGG GAG TCA TCA CCA AAT CCC ACA     8034
Cys Asp Thr Leu Leu Cys Asp Ile Gly Glu Ser Ser Pro Asn Pro Thr
                2635                2640                2645

GTG GAA GCA GGA CGA ACA CTC AGA GTC CTT AAC TTA GTA GAA AAT TGG     8082
Val Glu Ala Gly Arg Thr Leu Arg Val Leu Asn Leu Val Glu Asn Trp
            2650                2655                2660

TTG AAC AAC AAC ACT CAA TTT TGC ATA AAG GTT CTC AAC CCA TAT ATG     8130
Leu Asn Asn Asn Thr Gln Phe Cys Ile Lys Val Leu Asn Pro Tyr Met
        2665                2670                2675

CCC TCA GTC ATA GAA AAA ATG GAA GCA CTA CAA AGG AAA TAT GGA GGA     8178
Pro Ser Val Ile Glu Lys Met Glu Ala Leu Gln Arg Lys Tyr Gly Gly
   2680                2685                2690

GCC TTA GTG AGG AAT CCA CTC TCA CGA AAC TCC ACA CAT GAG ATG TAC     8226
Ala Leu Val Arg Asn Pro Leu Ser Arg Asn Ser Thr His Glu Met Tyr
2695                2700                2705                2710

TGG GTA TCC AAT GCT TCC GGG AAC ATA GTG TCA TCA GTG AAC ATG ATT     8274
Trp Val Ser Asn Ala Ser Gly Asn Ile Val Ser Ser Val Asn Met Ile
                2715                2720                2725

TCA AGG ATG TTG ATC AAC AGA TTT ACA ATG AGA TAC AAG AAA GCC ACT     8322
Ser Arg Met Leu Ile Asn Arg Phe Thr Met Arg Tyr Lys Lys Ala Thr
            2730                2735                2740

TAC GAG CCG GAT GTT GAC CTC GGA AGC GGA ACC CGT AAC ATC GGG ATT     8370
Tyr Glu Pro Asp Val Asp Leu Gly Ser Gly Thr Arg Asn Ile Gly Ile
        2745                2750                2755
```

```
GAA AGT GAG ATA CCA AAC CTA GAT ATA ATT GGG AAA AGA ATA GAA AAA     8418
Glu Ser Glu Ile Pro Asn Leu Asp Ile Ile Gly Lys Arg Ile Glu Lys
   2760                2765                2770

ATA AAG CAA GAG CAT GAA ACA TCA TGG CAC TAT GAC CAA GAC CAC CCA     8466
Ile Lys Gln Glu His Glu Thr Ser Trp His Tyr Asp Gln Asp His Pro
2775                2780                2785                2790

TAC AAA ACG TGG GCA TAC CAT GGT AGC TAT GAA ACA AAA CAG ACT GGA     8514
Tyr Lys Thr Trp Ala Tyr His Gly Ser Tyr Glu Thr Lys Gln Thr Gly
                2795                2800                2805

TCA GCA TCA TCC ATG GTC AAC GGA GTG GTC AGG CTG CTG ACA AAA CCT     8562
Ser Ala Ser Ser Met Val Asn Gly Val Val Arg Leu Leu Thr Lys Pro
           2810                2815                2820

TGG GAC GTC GTC CCC ATG GTG ACA CAG ATG GCA ATG ACA GAC ACG ACT     8610
Trp Asp Val Val Pro Met Val Thr Gln Met Ala Met Thr Asp Thr Thr
      2825                2830                2835

CCA TTT GGA CAA CAG CGC GTT TTT AAA GAG AAA GTG GAC ACG AGA ACC     8658
Pro Phe Gly Gln Gln Arg Val Phe Lys Glu Lys Val Asp Thr Arg Thr
   2840                2845                2850

CAA GAA CCG AAA GAA GGC ACG AAG AAA CTA ATG AAA ATA ACA GCA GAG     8706
Gln Glu Pro Lys Glu Gly Thr Lys Lys Leu Met Lys Ile Thr Ala Glu
2855                2860                2865                2870

TGG CTT TGG AAA GAA TTA GGG AAG AAA AAG ACA CCC AGG ATG TGC ACC     8754
Trp Leu Trp Lys Glu Leu Gly Lys Lys Lys Thr Pro Arg Met Cys Thr
                2875                2880                2885

AGA GAA GAA TTC ACA AGA AAG GTG AGA AGC AAT GCA GCC TTG GGG GCC     8802
Arg Glu Glu Phe Thr Arg Lys Val Arg Ser Asn Ala Ala Leu Gly Ala
           2890                2895                2900

ATA TTC ACT GAT GAG AAC AAG TGG AAG TCG GCA CGT GAG GCT GTT GAA     8850
Ile Phe Thr Asp Glu Asn Lys Trp Lys Ser Ala Arg Glu Ala Val Glu
      2905                2910                2915

GAT AGT AGG TTT TGG GAG CTG GTT GAC AAG GAA AGG AAT CTC CAT CTT     8898
Asp Ser Arg Phe Trp Glu Leu Val Asp Lys Glu Arg Asn Leu His Leu
   2920                2925                2930

GAA GGA AAG TGT GAA ACA TGT GTG TAC AAC ATG ATG GGA AAA AGA GAG     8946
Glu Gly Lys Cys Glu Thr Cys Val Tyr Asn Met Met Gly Lys Arg Glu
2935                2940                2945                2950

AAG AAG CTA GGG GAA TTC GGC AAG GCA AAA GGC AGC AGA GCC ATA TGG     8994
Lys Lys Leu Gly Glu Phe Gly Lys Ala Lys Gly Ser Arg Ala Ile Trp
                2955                2960                2965

TAC ATG TGG CTT GGA GCA CGC TTC TTA GAG TTT GAA GCC CTA GGA TTC     9042
Tyr Met Trp Leu Gly Ala Arg Phe Leu Glu Phe Glu Ala Leu Gly Phe
           2970                2975                2980

TTA AAT GAA GAT CAC TGG TTC TCC AGA GAG AAC TCC CTG AGT GGA GTG     9090
Leu Asn Glu Asp His Trp Phe Ser Arg Glu Asn Ser Leu Ser Gly Val
      2985                2990                2995
```

```
GAA GGA GAA GGG CTG CAC AAG CTA GGT TAC ATT CTA AGA GAC GTG AGC    9138
Glu Gly Glu Gly Leu His Lys Leu Gly Tyr Ile Leu Arg Asp Val Ser
   3000                3005                3010

AAG AAA GAG GGA GGA GCA ATG TAT GCC GAT GAC ACC GCA GGA TGG GAT    9186
Lys Lys Glu Gly Gly Ala Met Tyr Ala Asp Asp Thr Ala Gly Trp Asp
3015                3020                3025                3030

ACA AGA ATC ACA CTA GAA GAC KKA AAA AAT GAA GAA ATG GTA ACA AAC    9234
Thr Arg Ile Thr Leu Glu Asp Xaa Lys Asn Glu Glu Met Val Thr Asn
                3035                3040                3045

CAC ATG GAA GGA GAA CAC AAG AAA CTA GCC GAG GCC ATT TTC AAA CTA    9282
His Met Glu Gly Glu His Lys Lys Leu Ala Glu Ala Ile Phe Lys Leu
           3050                3055                3060

ACG TAC CAA AAC AAG GTG GTG CGT GTG CAA AGA CCA ACA CCA AGA GGC    9330
Thr Tyr Gln Asn Lys Val Val Arg Val Gln Arg Pro Thr Pro Arg Gly
      3065                3070                3075

ACA GTA ATG GAC ATC ATA TCG AGA AGA GAC CAA AGA GGT AGT GGA CAA    9378
Thr Val Met Asp Ile Ile Ser Arg Arg Asp Gln Arg Gly Ser Gly Gln
   3080                3085                3090

GTT GGC ACC TAT GGA CTC AAT ACT TTC ACC AAT ATG GAA GCC CAA CTA    9426
Val Gly Thr Tyr Gly Leu Asn Thr Phe Thr Asn Met Glu Ala Gln Leu
3095                3100                3105                3110

ATC AGA CAG ATG GAG GGA GAA GGA GTC TTT AAA AGC ATT CAG CAC CTA    9474
Ile Arg Gln Met Glu Gly Glu Gly Val Phe Lys Ser Ile Gln His Leu
                3115                3120                3125

ACA ATC ACA GAA GAA ATC GCT GTG CAA AAC TGG TTA GCA AGA GTG GGG    9522
Thr Ile Thr Glu Glu Ile Ala Val Gln Asn Trp Leu Ala Arg Val Gly
           3130                3135                3140

CGC GAA AGG TTA TCA AGA ATG GCC ATC AGT GGA GAT GAT TGT GTT GTG    9570
Arg Glu Arg Leu Ser Arg Met Ala Ile Ser Gly Asp Asp Cys Val Val
      3145                3150                3155

AAA CCT TTA GAT GAC AGG TTC GCA AGC GCT TTA ACA GCT CTA AAT GAC    9618
Lys Pro Leu Asp Asp Arg Phe Ala Ser Ala Leu Thr Ala Leu Asn Asp
   3160                3165                3170

ATG GGA AAG ATT AGG AAA GAC ATA CAA CAA TGG GAA CCT TCA AGA GGA    9666
Met Gly Lys Ile Arg Lys Asp Ile Gln Gln Trp Glu Pro Ser Arg Gly
3175                3180                3185                3190

TGG AAT GAT TGG ACA CAA GTG CCC TTC TGT TCA CAC CAT TTC CAT GAG    9714
Trp Asn Asp Trp Thr Gln Val Pro Phe Cys Ser His His Phe His Glu
                3195                3200                3205

TTA ATC ATG AAA GAC GGT CGC GTA CTC GTT GTT CCA TGT AGA AAC CAA    9762
Leu Ile Met Lys Asp Gly Arg Val Leu Val Val Pro Cys Arg Asn Gln
           3210                3215                3220

GAT GAA CTG ATT GGC AGA GCC CGA ATC TCC CAA GGA GCA GGG TGG TCT    9810
Asp Glu Leu Ile Gly Arg Ala Arg Ile Ser Gln Gly Ala Gly Trp Ser
      3225                3230                3235
```

```
TTG CGG GAG ACG GCC TGT TTG GGG AAG TCT TAC GCC CAA ATG TGG AGC   9858
Leu Arg Glu Thr Ala Cys Leu Gly Lys Ser Tyr Ala Gln Met Trp Ser
   3240                3245                3250

TTG ATG TAC TTC CAC AGA CGC GAC CTC AGG CTG GCG GCA AAT GCT ATT   9906
Leu Met Tyr Phe His Arg Arg Asp Leu Arg Leu Ala Ala Asn Ala Ile
3255                3260                3265                3270

TGC TCG GCA GTA CCA TCA CAT TGG GTT CCA ACA AGT CGA ACA ACC TGG   9954
Cys Ser Ala Val Pro Ser His Trp Val Pro Thr Ser Arg Thr Thr Trp
               3275                3280                3285

TCC ATA CAT GCT AAA CAT GAA TGG ATG ACA ACG GAA GAC ATG CTG ACA   10002
Ser Ile His Ala Lys His Glu Trp Met Thr Thr Glu Asp Met Leu Thr
          3290                3295                3300

GTC TGG AAC AGG GTG TGG ATT CAA GAA AAC CCA TGG ATG GAA GAC AAA   10050
Val Trp Asn Arg Val Trp Ile Gln Glu Asn Pro Trp Met Glu Asp Lys
     3305                3310                3315

ACT CCA GTG GAA TCA TGG GAG GAA ATC CCA TAC TTG GGG AAA AGA GAA   10098
Thr Pro Val Glu Ser Trp Glu Glu Ile Pro Tyr Leu Gly Lys Arg Glu
   3320                3325                3330

GAC CAA TGG TGC GGC TCA TTG ATT GGG TTA ACA AGC AGG GCC ACC TGG   10146
Asp Gln Trp Cys Gly Ser Leu Ile Gly Leu Thr Ser Arg Ala Thr Trp
3335                3340                3345                3350

GCA AAG AAC ATC CAA GCA GCA ATA AAT CAA GTT AGA TCC CTT ATA GGC   10194
Ala Lys Asn Ile Gln Ala Ala Ile Asn Gln Val Arg Ser Leu Ile Gly
               3355                3360                3365

AAT GAA GAA TAC ACA GAT TAC ATG CCA TCC ATG AAA AGA TTC AGA AGA   10242
Asn Glu Glu Tyr Thr Asp Tyr Met Pro Ser Met Lys Arg Phe Arg Arg
          3370                3375                3380

GAA GAG GAA GAA GCA GGA GTT CTG TGG TAGAAAGCAA AACTAACATG AAACAAGG 10297
Glu Glu Glu Glu Ala Gly Val Leu Trp
     3385                3390
CTAGAAGTCA GGTCGGATTA AGCCATAGTA CGGAAAAAAC TATGCTACCT GTGAGCCCCG  10357
TCCAAGGACG TTAAAAGAAG TCAGGCCATC ATAAATGCCA TAGCTTGAGT AAACTATGCA  10417
GCCTGTAGCT CCACCTGAGA AGGTGTAAAA AATCCGGGAG GCCACAAACC ATGGAAGCTG  10477
TACGCATGGC GTAGTGGACT AGCGGTTAGA GAGGACCCCT CCCTTACAAA TCGCAGCAAC  10537
AATGGGGGCC CAAGGCGAGA TGAAGCTGTA GTCTCGCTGG AAGGACTAGA GGTTAGAGGA  10597
GACCCCCCCG AAACAAAAAA CAGCATATTG ACGCTGGGAA AGACCAGAGA TCCTGCTGTC  10657
TCCTCAGCAT CATTCCAGGC ACAGAACGCC AGAAAATGGA ATGGTGCTGT TGAATCAACA  10717
GGTTCT                                                             10723
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 10723 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 97...10269
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
AGTTGTTAGT CTACGTGGAC CGACAAAGAC AGATTCTTTG AGGGAGCTAA GCTCAATGTA      60
GTTCTAACAG TTTTTTAATT AGAGAGCAGA TCTCTG ATG AAT AAC CAA CGG AAA      114
                                        Met Asn Asn Gln Arg Lys
                                         1               5

AAG GCG AAA AAC ACG CCT TTC AAT ATG CTG AAA CGC GAG AGA AAC CGC      162
Lys Ala Lys Asn Thr Pro Phe Asn Met Leu Lys Arg Glu Arg Asn Arg
            10                  15                  20

GTG TCG ACT GTG CAA CAG CTG ACA AAG AGA TTC TCA CTT GGA ATG CTG      210
Val Ser Thr Val Gln Gln Leu Thr Lys Arg Phe Ser Leu Gly Met Leu
        25                  30                  35

CAG GGA CGA GGA CCA TTA AAA CTG TTC ATG GCC CTG GTG GCG TTC CTT      258
Gln Gly Arg Gly Pro Leu Lys Leu Phe Met Ala Leu Val Ala Phe Leu
    40                  45                  50

CGT TTC CTA ACA ATC CCA CCA ACA GCA GGG ATA TTG AAG AGA TGG GGA      306
Arg Phe Leu Thr Ile Pro Pro Thr Ala Gly Ile Leu Lys Arg Trp Gly
55                  60                  65                  70

ACA ATT AAA AAA TCA AAA GCT ATT AAT GTT TTG AGA GGG TTC AGG AAA      354
Thr Ile Lys Lys Ser Lys Ala Ile Asn Val Leu Arg Gly Phe Arg Lys
                75                  80                  85

GAG ATT GGA AGG ATG CTG AAC ATC TTG AAT AGG AGA CGC AGA TCT GCA      402
Glu Ile Gly Arg Met Leu Asn Ile Leu Asn Arg Arg Arg Arg Ser Ala
            90                  95                  100

GGC ATG ATC ATT ATG CTG ATT CCA ACA GTG ATG GCG TTC CAT TTA ACC      450
Gly Met Ile Ile Met Leu Ile Pro Thr Val Met Ala Phe His Leu Thr
        105                 110                 115

ACA CGT AAC GGA GAA CCA CAC ATG ATC GTC AGC AGA CAA GAG AAA GGG      498
Thr Arg Asn Gly Glu Pro His Met Ile Val Ser Arg Gln Glu Lys Gly
    120                 125                 130

AAA AGT CTT CTG TTT AAA ACA GAG GTT GGC GTG AAC ATG TGT ACC CTC      546
Lys Ser Leu Leu Phe Lys Thr Glu Val Gly Val Asn Met Cys Thr Leu
135                 140                 145                 150

ATG GCC ATG GAC CTT GGT GAA TTG TGT GAA GAC ACA ATC ACG TAC AAG      594
Met Ala Met Asp Leu Gly Glu Leu Cys Glu Asp Thr Ile Thr Tyr Lys
                155                 160                 165

TGT CCC CTT CTC AGG CAG AAT GAG CCA GAA GAC ATA GAC TGT TGG TGC      642
Cys Pro Leu Leu Arg Gln Asn Glu Pro Glu Asp Ile Asp Cys Trp Cys
            170                 175                 180

NAC TCT ACG TCC ACG TGG GTA ACT TAT GGG ACG TGT ACC ACC ATG GGA      690
Xaa Ser Thr Ser Thr Trp Val Thr Tyr Gly Thr Cys Thr Thr Met Gly
        185                 190                 195
```

```
GAA CAT AGA AGA GAA AAA AGA TCA GTG GCA CTC GTT CCA CAT GTG GGA      738
Glu His Arg Arg Glu Lys Arg Ser Val Ala Leu Val Pro His Val Gly
    200                 205                 210

ATG GGA CTG GAG ACA CGA ACT GAA ACA TGG ATG TCA TCA GAA GGG GCC      786
Met Gly Leu Glu Thr Arg Thr Glu Thr Trp Met Ser Ser Glu Gly Ala
215                 220                 225                 230

TGG AAA CAT GTC CAG AGA ATT GAA ACT TGG ATC TTG AGA CAT CCA GGC      834
Trp Lys His Val Gln Arg Ile Glu Thr Trp Ile Leu Arg His Pro Gly
                235                 240                 245

TTC ACC ATG ATG GCA GCA ATC CTG GCA TAC ACC ATA GGA ACG ACA CAT      882
Phe Thr Met Met Ala Ala Ile Leu Ala Tyr Thr Ile Gly Thr Thr His
            250                 255                 260

TTC CAA AGA GCC CTG ATT TTC ATC TTA CTG ACA GCT GTC ACT CCT TCA      930
Phe Gln Arg Ala Leu Ile Phe Ile Leu Leu Thr Ala Val Thr Pro Ser
        265                 270                 275

ATG ACA ATG CGT TGC ATA GGA ATG TCA AAT AGA GAC TTT GTG GAA GGG      978
Met Thr Met Arg Cys Ile Gly Met Ser Asn Arg Asp Phe Val Glu Gly
    280                 285                 290

GTT TCA GGA GGA AGC TGG GTT GAC ATA GTC TTA GAA CAT GGA AGC TGT     1026
Val Ser Gly Gly Ser Trp Val Asp Ile Val Leu Glu His Gly Ser Cys
295                 300                 305                 310

GTG ACG ACG ATG GCA AAA AAC AAA CCA ACA TTG GAT TTT GAA CTG ATA     1074
Val Thr Thr Met Ala Lys Asn Lys Pro Thr Leu Asp Phe Glu Leu Ile
                315                 320                 325

AAA ACA GAA GCC AAA CAG CCT GCC ACC CTA AGG AAG TAC TGT ATA GAG     1122
Lys Thr Glu Ala Lys Gln Pro Ala Thr Leu Arg Lys Tyr Cys Ile Glu
            330                 335                 340

GCA AAG CTA ACC NAC ACA ACA ACA GAA TCT CGC TGC CCA ACA CAA GGG     1170
Ala Lys Leu Thr Xaa Thr Thr Thr Glu Ser Arg Cys Pro Thr Gln Gly
        345                 350                 355

GAA CCC AGC CTA AAT GAA GAG CAG GAC AAA AGG TTC GTC TGC AAA CAC     1218
Glu Pro Ser Leu Asn Glu Glu Gln Asp Lys Arg Phe Val Cys Lys His
    360                 365                 370

TCC ATG GTA GAC AGA GGA TGG GGA AAT GGA TGT GGA CTA TTT GGA AAG     1266
Ser Met Val Asp Arg Gly Trp Gly Asn Gly Cys Gly Leu Phe Gly Lys
375                 380                 385                 390

GGA GGC ATT GTG ACC TGT GCT ATG TTC AGA TGC AAA AAG AAC ATG GAA     1314
Gly Gly Ile Val Thr Cys Ala Met Phe Arg Cys Lys Lys Asn Met Glu
                395                 400                 405

GGA AAA GTT GTG CAA CCA GAA AAC TTG GAA TAC ACC ATT GTG ATA ACA     1362
Gly Lys Val Val Gln Pro Glu Asn Leu Glu Tyr Thr Ile Val Ile Thr
            410                 415                 420

CCT CAC TCA GGG GAA GAG CAT GCA GTC GGA NAT GAC ACA GGA AAA CAT     1410
Pro His Ser Gly Glu Glu His Ala Val Gly Xaa Asp Thr Gly Lys His
        425                 430                 435
```

```
GGC AAG GAA ATC AAA ATA ACA CCA CAG AGT TCC ATC ACA GAA GCA GAA     1458
Gly Lys Glu Ile Lys Ile Thr Pro Gln Ser Ser Ile Thr Glu Ala Glu
    440                 445                 450

TTG ACA GGT TAT GGC ACT GTC ACA ATG GAG TGC TCT CCA AGA ACG GGC     1506
Leu Thr Gly Tyr Gly Thr Val Thr Met Glu Cys Ser Pro Arg Thr Gly
455                 460                 465                 470

CTC GAC TTC AAT GAG ATG GTG TTG CTG CAG ATG GAA AAT AAA GCT TGG     1554
Leu Asp Phe Asn Glu Met Val Leu Leu Gln Met Glu Asn Lys Ala Trp
                475                 480                 485

CTG GTG CAC AGG CAA TGG TTC CTA GAC CTG CCG TTA CCA TGG TTG CCC     1602
Leu Val His Arg Gln Trp Phe Leu Asp Leu Pro Leu Pro Trp Leu Pro
            490                 495                 500

GGA GCG GAC ACA CAA GGG TCA AAT TGG ATA CAG AAA GAG ACA TTG GTC     1650
Gly Ala Asp Thr Gln Gly Ser Asn Trp Ile Gln Lys Glu Thr Leu Val
        505                 510                 515

ACT TTC AAA AAT CCC CAT GCG AAG AAA CAG GAT GTT GTT GTT TTA GGA     1698
Thr Phe Lys Asn Pro His Ala Lys Lys Gln Asp Val Val Val Leu Gly
    520                 525                 530

TCC CAA GAA GGG GCC ATG CAC ACA GCA CTT ACA GGG GCC ACA GAA ATC     1746
Ser Gln Glu Gly Ala Met His Thr Ala Leu Thr Gly Ala Thr Glu Ile
535                 540                 545                 550

CAA ATG TCA TCA GGA AAC TTA CTC TTC ACA GGA CAT CTC AAG TGC AGG     1794
Gln Met Ser Ser Gly Asn Leu Leu Phe Thr Gly His Leu Lys Cys Arg
                555                 560                 565

CTG AGA ATG GAC AAG CTA CAG CTC AAA GGA ATG TCA TAC TCT ATG TGC     1842
Leu Arg Met Asp Lys Leu Gln Leu Lys Gly Met Ser Tyr Ser Met Cys
            570                 575                 580

ACA GGA AAG TTT AAA GTT GTG AAG GAA ATA GCA GAA ACA CAA CAT GGA     1890
Thr Gly Lys Phe Lys Val Val Lys Glu Ile Ala Glu Thr Gln His Gly
        585                 590                 595

ACA ATA GTT ATC AGA GTG CAA TAT GAA GGG GAC GGC TCT CCA TGC AAG     1938
Thr Ile Val Ile Arg Val Gln Tyr Glu Gly Asp Gly Ser Pro Cys Lys
    600                 605                 610

ATC CCT TTT GAG ATA ATG GAT TTG GAA AAA AGA CAT GTC TTA GGT CGC     1986
Ile Pro Phe Glu Ile Met Asp Leu Glu Lys Arg His Val Leu Gly Arg
615                 620                 625                 630

CTG ATT ACA GTC AAC CCA ATT GTG ACA GAA AAA GAT AGC CCA GTC AAC     2034
Leu Ile Thr Val Asn Pro Ile Val Thr Glu Lys Asp Ser Pro Val Asn
                635                 640                 645

ATA GAA GCA GAA CCT CCA TTT GGA GAC AGC TAC ATC ATC ATA GGA GTA     2082
Ile Glu Ala Glu Pro Pro Phe Gly Asp Ser Tyr Ile Ile Ile Gly Val
            650                 655                 660

GAG CCG GGA CAA CTG AAG CTC AAC TGG TTT AAG AAA GGA AGT TCT ATC     2130
Glu Pro Gly Gln Leu Lys Leu Asn Trp Phe Lys Lys Gly Ser Ser Ile
        665                 670                 675
```

```
GGC CAA ATG TTT GAG ACA ACA ATG AGG GGG GCG AAG AGA ATG GCC ATT     2178
Gly Gln Met Phe Glu Thr Thr Met Arg Gly Ala Lys Arg Met Ala Ile
    680                 685                 690

TTA GGT GAC ACA GCC TGG GAT TTT GGA TCC TTG GGA GGA GTG TTT ACA     2226
Leu Gly Asp Thr Ala Trp Asp Phe Gly Ser Leu Gly Gly Val Phe Thr
695                 700                 705                 710

TCT ATA GGA AAG GCT CTC CAC CAA GTC TTT GGA GCA ATC TAT GGA GCT     2274
Ser Ile Gly Lys Ala Leu His Gln Val Phe Gly Ala Ile Tyr Gly Ala
                715                 720                 725

GCC TTC AGT GGG GTT TCA TGG ACT ATG AAA ATC CTC ATA GGA GTC ATT     2322
Ala Phe Ser Gly Val Ser Trp Thr Met Lys Ile Leu Ile Gly Val Ile
            730                 735                 740

ATC ACA TGG ATA GGA ATG AAT TCA CGC AGC ACC TCA CTG TCT GTG ACA     2370
Ile Thr Trp Ile Gly Met Asn Ser Arg Ser Thr Ser Leu Ser Val Thr
        745                 750                 755

CTA GTA TTG GTG GGA ATT GTG ACA CTG TAT TTG GGA GTC ATG GTG CAG     2418
Leu Val Leu Val Gly Ile Val Thr Leu Tyr Leu Gly Val Met Val Gln
    760                 765                 770

GCC GAT AGT GGT TGC GTT GTG AGC TGG AAA AAC AAA GAA CTG AAA TGT     2466
Ala Asp Ser Gly Cys Val Val Ser Trp Lys Asn Lys Glu Leu Lys Cys
775                 780                 785                 790

GGC AGT GGG ATT TTC ATC ACA GAC AAC GTG CAC ACA TGG ACA GAA CAA     2514
Gly Ser Gly Ile Phe Ile Thr Asp Asn Val His Thr Trp Thr Glu Gln
                795                 800                 805

TAC AAG TTC CAA CCA GAA TCC CCT TCA AAA CTA GCT TCA GCT ATC CAG     2562
Tyr Lys Phe Gln Pro Glu Ser Pro Ser Lys Leu Ala Ser Ala Ile Gln
            810                 815                 820

AAA GCC CAT GAA GAG GAC ATT TGT GGA ATC CGC TCA GTA ACA AGA CTG     2610
Lys Ala His Glu Glu Asp Ile Cys Gly Ile Arg Ser Val Thr Arg Leu
        825                 830                 835

GAG AAT CTG ATG TGG AAA CAA ATA ACA CCA GAA TTG AAT CAC ATT CTA     2658
Glu Asn Leu Met Trp Lys Gln Ile Thr Pro Glu Leu Asn His Ile Leu
    840                 845                 850

TCA GAA AAT GAG GTG AAG TTA ACT ATT ATG ACA GGA GAC ATC AAA GGA     2706
Ser Glu Asn Glu Val Lys Leu Thr Ile Met Thr Gly Asp Ile Lys Gly
855                 860                 865                 870

ATC ATG CAG GCA GGA AAA CGA TCT CTG CGG CCT CAG CCC ACT GAG CTG     2754
Ile Met Gln Ala Gly Lys Arg Ser Leu Arg Pro Gln Pro Thr Glu Leu
                875                 880                 885

AAG TAT TCA TGG AAA ACA TGG GGC AAA GCA AAA ATG CTC TCT ACA GAG     2802
Lys Tyr Ser Trp Lys Thr Trp Gly Lys Ala Lys Met Leu Ser Thr Glu
            890                 895                 900

TCT CAT NAC CAG ACC TTT CTC ATT GAT GGC CCC GAA ACA GCA GAA TGC     2850
Ser His Xaa Gln Thr Phe Leu Ile Asp Gly Pro Glu Thr Ala Glu Cys

        905                 910                 915
```

```
CCC AAC ACA AAT AGA GCT TGG AAT TCG TTG GAA GTT GAA GAC TAT GGC     2898
Pro Asn Thr Asn Arg Ala Trp Asn Ser Leu Glu Val Glu Asp Tyr Gly
    920                 925                 930

TTT GGA GTA TTC ACC ACC AAT ATA TGG CTA AAA TTG AAA GAA AAA CAG     2946
Phe Gly Val Phe Thr Thr Asn Ile Trp Leu Lys Leu Lys Glu Lys Gln
935                 940                 945                 950

GAT GTA TTC TGC GAC TCA AAA CTC ATG TCA GCG GCC ATA AAA GAC AAC     2994
Asp Val Phe Cys Asp Ser Lys Leu Met Ser Ala Ala Ile Lys Asp Asn
                955                 960                 965

AGA GCC GTC CAT GCC GAT ATG GGT TAT TGG ATA GAA AGT GCA CTC NAT     3042
Arg Ala Val His Ala Asp Met Gly Tyr Trp Ile Glu Ser Ala Leu Xaa
            970                 975                 980

GAC ACA TGG AAG ATA GAG AAA GCC TCT TTC ATT GAA GTT AAA AAC TGC     3090
Asp Thr Trp Lys Ile Glu Lys Ala Ser Phe Ile Glu Val Lys Asn Cys
        985                 990                 995

CAC TGG CCA AAA TCA CAC ACC CTC TGG AGC AAT GGA GTG CTA GAA AGT     3138
His Trp Pro Lys Ser His Thr Leu Trp Ser Asn Gly Val Leu Glu Ser
   1000                1005                1010

GAG ATG ATA ATT CCA AAG AAT CTC GCT GGA CCA GTG TCT CAA CAC AAC     3186
Glu Met Ile Ile Pro Lys Asn Leu Ala Gly Pro Val Ser Gln His Asn
1015                1020                1025                1030

TAT AGA CCA GGC TAC CAT ACA CAA ATA ACA GGA CCA TGG CAT CTA GGT     3234
Tyr Arg Pro Gly Tyr His Thr Gln Ile Thr Gly Pro Trp His Leu Gly
                1035                1040                1045

AAG CTT GAG ATG GAC TTT GAT TTC TGT GAT GGA ACA ACA GTG GTA GTG     3282
Lys Leu Glu Met Asp Phe Asp Phe Cys Asp Gly Thr Thr Val Val Val
            1050                1055                1060

ACT GAG GAC TGC GGA AAT AGA GGA CCC TCT TTG AGA ACA ACC ACT GCC     3330
Thr Glu Asp Cys Gly Asn Arg Gly Pro Ser Leu Arg Thr Thr Thr Ala
        1065                1070                1075

TCT GGA AAA CTC ATA ACA GAA TGG TGC TGC CGA TCT TGC ACA TTA CCA     3378
Ser Gly Lys Leu Ile Thr Glu Trp Cys Cys Arg Ser Cys Thr Leu Pro
   1080                1085                1090

CCG CTA AGA TAC AGA GGT GAG GAT GGG TGC TGG TAC GGG ATG GAA ATC     3426
Pro Leu Arg Tyr Arg Gly Glu Asp Gly Cys Trp Tyr Gly Met Glu Ile
1095                1100                1105                1110

AGA CCA TTG AAG GAG AAA GAA GAG AAT TTG GTC AAC TCC TTG GTC ACA     3474
Arg Pro Leu Lys Glu Lys Glu Glu Asn Leu Val Asn Ser Leu Val Thr
                1115                1120                1125

GCT GGA CAT GGG CAG GTC GAC AAC TTT TCA CTA GGA GTC TTG GGA ATG     3522
Ala Gly His Gly Gln Val Asp Asn Phe Ser Leu Gly Val Leu Gly Met
            1130                1135                1140

GCA TTG TTC CTG GAG GAA ATG CTT AGG ACC CGA GTA GGA ACG AAA CAT     3570
Ala Leu Phe Leu Glu Glu Met Leu Arg Thr Arg Val Gly Thr Lys His
        1145                1150                1155
```

```
GCA ATA CTA CTA GTT GCA GTT TCT TTT GTG ACA TTG ATC ACA GGG AAC     3618
Ala Ile Leu Leu Val Ala Val Ser Phe Val Thr Leu Ile Thr Gly Asn
   1160                1165                1170

ATG TCC TTT AGA GAC CTG GGA AGA GTG ATG GTT ATG GTA GGC GCC ACT     3666
Met Ser Phe Arg Asp Leu Gly Arg Val Met Val Met Val Gly Ala Thr
1175                1180                1185                1190

ATG ACG GAT GAC ATA GGT ATG GGC GTG ACT TAT CTT GCC CTA CTA GCA     3714
Met Thr Asp Asp Ile Gly Met Gly Val Thr Tyr Leu Ala Leu Leu Ala
                1195                1200                1205

GCC TTC AAA GTC AGA CCA ACT TTT GCA GCT GGA CTA CTC TTG AGA AAG     3762
Ala Phe Lys Val Arg Pro Thr Phe Ala Ala Gly Leu Leu Leu Arg Lys
            1210                1215                1220

CTG ACC TCC AAG GAA TTG ATG ATG ACT ACT ATA GGA ATT GTA CTC CTC     3810
Leu Thr Ser Lys Glu Leu Met Met Thr Thr Ile Gly Ile Val Leu Leu
        1225                1230                1235

TCC CAG AGC ACC ATA CCA GAG ACC ATT CTT GAG TTG ACT GAT GCG TTA     3858
Ser Gln Ser Thr Ile Pro Glu Thr Ile Leu Glu Leu Thr Asp Ala Leu
   1240                1245                1250

GCC TTA GGC ATG ATG GTC CTC AAA ATG GTG AGA AAT ATG GAA AAG TAT     3906
Ala Leu Gly Met Met Val Leu Lys Met Val Arg Asn Met Glu Lys Tyr
1255                1260                1265                1270

CAA TTG GCA GTG ACT ATC ATG GCT ATC TTG TGC GTC CCA AAC GCA GTG     3954
Gln Leu Ala Val Thr Ile Met Ala Ile Leu Cys Val Pro Asn Ala Val
                1275                1280                1285

ATA TTA CAA AAC GCA TGG AAA GTG AGT TGC ACA ATA TTG GCA GTG GTG     4002
Ile Leu Gln Asn Ala Trp Lys Val Ser Cys Thr Ile Leu Ala Val Val
            1290                1295                1300

TCC GTT TCC CCA CTG TTC TTA ACA TCC TCA CAG CAA AAA ACA GAT TGG     4050
Ser Val Ser Pro Leu Phe Leu Thr Ser Ser Gln Gln Lys Thr Asp Trp
        1305                1310                1315

ATA CCA TTA GCA TTG ACG ATC AAA GGT CTC AAT CCA ACA GCT ATT TTT     4098
Ile Pro Leu Ala Leu Thr Ile Lys Gly Leu Asn Pro Thr Ala Ile Phe
   1320                1325                1330

CTA ACA ACC CTC TCA AGA ACC AGC AAG AAA AGG AGC TGG CCA TTA AAT     4146
Leu Thr Thr Leu Ser Arg Thr Ser Lys Lys Arg Ser Trp Pro Leu Asn
1335                1340                1345                1350

GAG GCT ATC ATG GCA GTC GGG ATG GTG AGC ATT TTA GCC AGT TCT CTC     4194
Glu Ala Ile Met Ala Val Gly Met Val Ser Ile Leu Ala Ser Ser Leu
                1355                1360                1365

CTA AAA AAT GAT ATT CCC ATG ACA GGA CCA TTA GTG GCT GGA GGG CTC     4242
Leu Lys Asn Asp Ile Pro Met Thr Gly Pro Leu Val Ala Gly Gly Leu
            1370                1375                1380

CTC ACT GTG TGC TAC GTG CTC ACT GGA CGA TCG GCC GAT TTG GAA CTG     4290
Leu Thr Val Cys Tyr Val Leu Thr Gly Arg Ser Ala Asp Leu Glu Leu
        1385                1390                1395
```

```
GAG AGA GCA GCC GAT GTC AAA TGG GAA GAC CAG GCA GAG ATA TCA GGA     4338
Glu Arg Ala Ala Asp Val Lys Trp Glu Asp Gln Ala Glu Ile Ser Gly
   1400                1405                1410

AGC AGT CCA ATC CTG TCA ATA ACA ATA TCA GAA GAT GGT AGC ATG TCG     4386
Ser Ser Pro Ile Leu Ser Ile Thr Ile Ser Glu Asp Gly Ser Met Ser
1415                1420                1425                1430

ATA AAA AAT GAA GAG GAA GAA CAA ACA CTG ACC ATA CTC ATT AGA ACA     4434
Ile Lys Asn Glu Glu Glu Glu Gln Thr Leu Thr Ile Leu Ile Arg Thr
                1435                1440                1445

GGA TTG CTG GTG ATC TCA GGA CTT TTT CCT GTA TCA ATA CCA ATC ACG     4482
Gly Leu Leu Val Ile Ser Gly Leu Phe Pro Val Ser Ile Pro Ile Thr
            1450                1455                1460

GCA GCA GCA TGG TAC CTG TGG GAA GTG AAG AAA CAA CGG GCC GGA GTA     4530
Ala Ala Ala Trp Tyr Leu Trp Glu Val Lys Lys Gln Arg Ala Gly Val
        1465                1470                1475

TTG TGG GAT GTT CCT TCA CCC CCA CCC ATG GGA AAG GCT GAA CTG GAA     4578
Leu Trp Asp Val Pro Ser Pro Pro Pro Met Gly Lys Ala Glu Leu Glu
   1480                1485                1490

GAT GGA GCC TAT AGA ATT AAG CAA AAA GGG ATT CTT GGA TAT TCC CAG     4626
Asp Gly Ala Tyr Arg Ile Lys Gln Lys Gly Ile Leu Gly Tyr Ser Gln
1495                1500                1505                1510

ATC GGA GCC GGA GTT TAC AAA GAA GGA ACA TTC CAT ACA ATG TGG CAT     4674
Ile Gly Ala Gly Val Tyr Lys Glu Gly Thr Phe His Thr Met Trp His
                1515                1520                1525

GTC ACA CGT GGC GCT GTT CTA ATG CAT AAA GGA AAG AGG ATT GAA CCA     4722
Val Thr Arg Gly Ala Val Leu Met His Lys Gly Lys Arg Ile Glu Pro
            1530                1535                1540

TCA TGG GCG GAC GTC AAG AAA GAC CTA ATA TCA TAT GGA GGA GGC TGG     4770
Ser Trp Ala Asp Val Lys Lys Asp Leu Ile Ser Tyr Gly Gly Gly Trp
        1545                1550                1555

AAG TTA GAA GGA GAA TGG AAG GAA GGA GAA GAA GTC CAG GTA TTG GCA     4818
Lys Leu Glu Gly Glu Trp Lys Glu Gly Glu Glu Val Gln Val Leu Ala
   1560                1565                1570

CTG GAG CCT GGA AAA AAT CCA AGA GCC GTC CAA ACG AAA CCT GGT CTT     4866
Leu Glu Pro Gly Lys Asn Pro Arg Ala Val Gln Thr Lys Pro Gly Leu
1575                1580                1585                1590

TTC AAA ACC AAC GCC GGA ACA ATA GGT GCT GTA TCT CTG GAC TTT TCT     4914
Phe Lys Thr Asn Ala Gly Thr Ile Gly Ala Val Ser Leu Asp Phe Ser
                1595                1600                1605

CCT GGA ACG TCA GGA TCT CCA ATT ATC GAC AAA AAA GGA AAA GTT GTG     4962
Pro Gly Thr Ser Gly Ser Pro Ile Ile Asp Lys Lys Gly Lys Val Val
            1610                1615                1620

GGT CTT TAT GGT AAT GGT GTT GTT ACA AGG AGT GGA GCA TAT GTG AGT     5010
Gly Leu Tyr Gly Asn Gly Val Val Thr Arg Ser Gly Ala Tyr Val Ser
        1625                1630                1635
```

```
GCT ATA GCC CAG ACT GAA AAA AGC ATT GAA GAC AAC CCA GAG ATC GAA    5058
Ala Ile Ala Gln Thr Glu Lys Ser Ile Glu Asp Asn Pro Glu Ile Glu
   1640                1645                1650

GAT GAC ATT TTC CGA AAG AGA AGA CTG ACC ATC ATG GAC CTC CAC CCA    5106
Asp Asp Ile Phe Arg Lys Arg Arg Leu Thr Ile Met Asp Leu His Pro
1655                1660                1665                1670

GGA GCG GGA AAG ACG AAG AGA TAC CTT CCG GCC ATA GTC AGA GAA GCT    5154
Gly Ala Gly Lys Thr Lys Arg Tyr Leu Pro Ala Ile Val Arg Glu Ala
                1675                1680                1685

ATA AAA CGG GGT TTG AGA ACA TTA ATC TTG GCC CCC ACT AGA GTT GTG    5202
Ile Lys Arg Gly Leu Arg Thr Leu Ile Leu Ala Pro Thr Arg Val Val
            1690                1695                1700

GCA GCT GAA ATG GAG GAA GCC CTT AGA GGA CTT CCA ATA AGA TAC CAG    5250
Ala Ala Glu Met Glu Glu Ala Leu Arg Gly Leu Pro Ile Arg Tyr Gln
        1705                1710                1715

ACC CCA GCC ATC AGA GCT GAG CAC ACC GGG CGG GAG ATT GTG GAC CTA    5298
Thr Pro Ala Ile Arg Ala Glu His Thr Gly Arg Glu Ile Val Asp Leu
   1720                1725                1730

ATG TGT CAT GCC ACA TTT ACC ATG AGG CTG CTA TCA CCA GTT AGA GTG    5346
Met Cys His Ala Thr Phe Thr Met Arg Leu Leu Ser Pro Val Arg Val
1735                1740                1745                1750

CCA AAC TAC AAC CTG ATT ATC ATG GAC GAA GCC CAT TTC ACA GAC CCA    5394
Pro Asn Tyr Asn Leu Ile Ile Met Asp Glu Ala His Phe Thr Asp Pro
                1755                1760                1765

GCA AGT ATA GCA GCT AGA GGA TAC ATC TCA ACT CGA GTG GAG ATG GGT    5442
Ala Ser Ile Ala Ala Arg Gly Tyr Ile Ser Thr Arg Val Glu Met Gly
            1770                1775                1780

GAG GCA GCT GGG ATT TTT ATG ACA GCC ACT CCC CCG GGA AGC AGA GAC    5490
Glu Ala Ala Gly Ile Phe Met Thr Ala Thr Pro Pro Gly Ser Arg Asp
        1785                1790                1795

CCA TTT CCT CAG AGC AAT GCA CCA ATC ATA GAT GAA GAA AGA GAA ATC    5538
Pro Phe Pro Gln Ser Asn Ala Pro Ile Ile Asp Glu Glu Arg Glu Ile
   1800                1805                1810

CCT GAA CGT TCG TGG AAT TCC GGA CAT GAA TGG GTC ACG GAT TTT AAA    5586
Pro Glu Arg Ser Trp Asn Ser Gly His Glu Trp Val Thr Asp Phe Lys
1815                1820                1825                1830

GGG AAG ACT GTT TGG TTC GTT CCA AGT ATA AAA GCA GGA AAT GAT ATA    5634
Gly Lys Thr Val Trp Phe Val Pro Ser Ile Lys Ala Gly Asn Asp Ile
                1835                1840                1845

GCA GCT TGC CTG AGG AAA AAT GGA AAG AAA GTG ATA CAA CTC AGT AGG    5682
Ala Ala Cys Leu Arg Lys Asn Gly Lys Lys Val Ile Gln Leu Ser Arg
            1850                1855                1860

AAG ACC TTT GAT TCT GAG TAT GTC AAG ACT AGA ACC AAT GAT TGG GAC    5730
Lys Thr Phe Asp Ser Glu Tyr Val Lys Thr Arg Thr Asn Asp Trp Asp
        1865                1870                1875
```

```
TTC GTG GTT ACA ACT GAC ATT TCA GAA ATG GGT GCC AAT TTC AAG GCT     5778
Phe Val Val Thr Thr Asp Ile Ser Glu Met Gly Ala Asn Phe Lys Ala
   1880                1885                1890

GAG AGG GTT ATA GAC CCC AGA CGC TGC ATG AAA CCA GTC ATA CTA ACA     5826
Glu Arg Val Ile Asp Pro Arg Arg Cys Met Lys Pro Val Ile Leu Thr
1895                1900                1905                1910

GAT GGT GAA GAG CGG GTG ATT CTG GCA GGA CCT ATG CCA GTG ACC CAC     5874
Asp Gly Glu Glu Arg Val Ile Leu Ala Gly Pro Met Pro Val Thr His
                1915                1920                1925

TCT AGT GCA GCA CAA AGA AGA GGG AGA ATA GGA AGA AAT CCA AAA AAT     5922
Ser Ser Ala Ala Gln Arg Arg Gly Arg Ile Gly Arg Asn Pro Lys Asn
            1930                1935                1940

GAG AAT GAC CAG TAC ATA TAC ATG GGG GAA CCT CTG GAA AAT GAT GAA     5970
Glu Asn Asp Gln Tyr Ile Tyr Met Gly Glu Pro Leu Glu Asn Asp Glu
        1945                1950                1955

GAC TGT GCA CAC TGG AAA GAA GCT AAA ATG CTC CTA GAT AAC ATC AAC     6018
Asp Cys Ala His Trp Lys Glu Ala Lys Met Leu Leu Asp Asn Ile Asn
   1960                1965                1970

ACG CCA GAA GGA ATC ATT CCT AGC ATG TTC GAA CCA GAG CGT GAA AAG     6066
Thr Pro Glu Gly Ile Ile Pro Ser Met Phe Glu Pro Glu Arg Glu Lys
1975                1980                1985                1990

GTG GAT GCC ATT GAT GGC GAA TAC CGC TTG AGA GGA GAA GCA AGG AAA     6114
Val Asp Ala Ile Asp Gly Glu Tyr Arg Leu Arg Gly Glu Ala Arg Lys
                1995                2000                2005

ACC TTT GTA GAC TTA ATG AGA AGA GGA GAC CTA CCA GTC TGG TTG GCC     6162
Thr Phe Val Asp Leu Met Arg Arg Gly Asp Leu Pro Val Trp Leu Ala
            2010                2015                2020

TAC AGA GTG GCA GCT GAA GGC ATC AAC TAC GCA GAC AGA AGG TGG TGT     6210
Tyr Arg Val Ala Ala Glu Gly Ile Asn Tyr Ala Asp Arg Arg Trp Cys
        2025                2030                2035

TTT GAT GGA GTC AAG AAC AAC CAA ATC CTA GAA GAA AAC GTG GAA GTT     6258
Phe Asp Gly Val Lys Asn Asn Gln Ile Leu Glu Glu Asn Val Glu Val
   2040                2045                2050

GAA ATC TGG ACA AAA GAA GGG GAA AGG AAG AAA TTG AAA CCC AGA TGG     6306
Glu Ile Trp Thr Lys Glu Gly Glu Arg Lys Lys Leu Lys Pro Arg Trp
2055                2060                2065                2070

TTG GAT GCT AGG ATC TAT TCT GAC CCA CTG GCG CTA AAA GAA TTT AAG     6354
Leu Asp Ala Arg Ile Tyr Ser Asp Pro Leu Ala Leu Lys Glu Phe Lys
                2075                2080                2085

GAA TTT GCA GCC GGA AGA AAG TCT CTG ACC CTG AAC CTA ATC ACA GAA     6402
Glu Phe Ala Ala Gly Arg Lys Ser Leu Thr Leu Asn Leu Ile Thr Glu
            2090                2095                2100

ATG GGT AGG CTC CCA ACC TTC ATG ACT CAG AAG GCA AGA GAC GCA CTG     6450
Met Gly Arg Leu Pro Thr Phe Met Thr Gln Lys Ala Arg Asp Ala Leu
        2105                2110                2115
```

```
GAC AAC TTA GCA GTG CTG CAC ACG GCT GAG GCA GGT GGA AGG GCG TAC     6498
Asp Asn Leu Ala Val Leu His Thr Ala Glu Ala Gly Gly Arg Ala Tyr
   2120                2125                2130

AAC CAT GCT CTC AGT GAA CTG CCG GAG ACC CTG GAG ACA TTG CTT TTA     6546
Asn His Ala Leu Ser Glu Leu Pro Glu Thr Leu Glu Thr Leu Leu Leu
2135                2140                2145                2150

CTG ACA CTT CTG GCT ACA GTC ACG GGA GGG ATC TTT TTA TTC TTG ATG     6594
Leu Thr Leu Leu Ala Thr Val Thr Gly Gly Ile Phe Leu Phe Leu Met
                2155                2160                2165

AGC GCA AGG GGC ATA GGG AAG ATG ACC CTG GGA ATG TGC TGC ATA ATC     6642
Ser Ala Arg Gly Ile Gly Lys Met Thr Leu Gly Met Cys Cys Ile Ile
           2170                2175                2180

ACG GCT AGC ATC CTC CTA TGG TAC GCA CAA ATA CAG CCA CAC TGG ATA     6690
Thr Ala Ser Ile Leu Leu Trp Tyr Ala Gln Ile Gln Pro His Trp Ile
      2185                2190                2195

GCA GCT TCA ATA ATA CTG GAG TTT TTT CTC ATA GTT TTG CTT ATT CCA     6738
Ala Ala Ser Ile Ile Leu Glu Phe Phe Leu Ile Val Leu Leu Ile Pro
   2200                2205                2210

GAA CCT GAA AAA CAG AGA ACA CCC CAA GAC AAC CAA CTG ACC TAC GTT     6786
Glu Pro Glu Lys Gln Arg Thr Pro Gln Asp Asn Gln Leu Thr Tyr Val
2215                2220                2225                2230

GTC ATA GCC ATC CTC ACA GTG GTG GCC GCA ACC ATG GCA AAC GAG ATG     6834
Val Ile Ala Ile Leu Thr Val Val Ala Ala Thr Met Ala Asn Glu Met
                2235                2240                2245

GGT TTC CTA GAA AAA ACG AAG AAA GAT CTC GGA TTG GGA AGC ATT GCA     6882
Gly Phe Leu Glu Lys Thr Lys Lys Asp Leu Gly Leu Gly Ser Ile Ala
           2250                2255                2260

ACC CAG CAA CCC GAG AGC AAC ATC CTG GAC ATA GAT CTA CGT CCT GCA     6930
Thr Gln Gln Pro Glu Ser Asn Ile Leu Asp Ile Asp Leu Arg Pro Ala
      2265                2270                2275

TCA GCA TGG ACG CTG TAT GCC GTG GCC ACA ACA TTT GTT ACA CCA ATG     6978
Ser Ala Trp Thr Leu Tyr Ala Val Ala Thr Thr Phe Val Thr Pro Met
   2280                2285                2290

TTG AGA CAT AGC ATT GAA AAT TCC TCA GTG AAT GTG TCC CTA ACA GCT     7026
Leu Arg His Ser Ile Glu Asn Ser Ser Val Asn Val Ser Leu Thr Ala
2295                2300                2305                2310

ATA GCC AAC CAA GCC ACA GTG TTA ATG GGT CTC GGG AAA GGA TGG CCA     7074
Ile Ala Asn Gln Ala Thr Val Leu Met Gly Leu Gly Lys Gly Trp Pro
                2315                2320                2325

TTG TCA AAG ATG GAC ATC GGA GTT CCC CTT CTC GCC ATT GGA TGC TAC     7122
Leu Ser Lys Met Asp Ile Gly Val Pro Leu Leu Ala Ile Gly Cys Tyr
           2330                2335                2340

TCA CAA GTC AAC CCC ATA ACT CTC ACA GCA GCT CTT TTC TTA TTG GTA     7170
Ser Gln Val Asn Pro Ile Thr Leu Thr Ala Ala Leu Phe Leu Leu Val
      2345                2350                2355
```

```
GCA CAT TAT GCC ATC ATA GGG CCA GGA CTC CAA GCA AAA GCA ACC AGA     7218
Ala His Tyr Ala Ile Ile Gly Pro Gly Leu Gln Ala Lys Ala Thr Arg
   2360                2365                2370

GAA GCT CAG AAA AGA GCA GCG GCG GGC ATC ATG AAA AAC CCA ACT GTC     7266
Glu Ala Gln Lys Arg Ala Ala Ala Gly Ile Met Lys Asn Pro Thr Val
2375                2380                2385                2390

GAT GGA ATA ACA GTG ATT GAC CTA GAT CCA ATA CCT TAT GAT CCA AAG     7314
Asp Gly Ile Thr Val Ile Asp Leu Asp Pro Ile Pro Tyr Asp Pro Lys
                2395                2400                2405

TTT GAA AAG CAG TTG GGA CAA GTA ATG CTC CTA GTC CTC TGC GTG ACT     7362
Phe Glu Lys Gln Leu Gly Gln Val Met Leu Leu Val Leu Cys Val Thr
           2410                2415                2420

CAA GTA TTG ATG ATG AGG ACT ACA TGG GCT CTG TGT GAG GCT TTA ACC     7410
Gln Val Leu Met Met Arg Thr Thr Trp Ala Leu Cys Glu Ala Leu Thr
      2425                2430                2435

TTA GCT ACC GGG CCC ATC TCC ACA TTG TGG GAA GGA AAT CCA GGG AGG     7458
Leu Ala Thr Gly Pro Ile Ser Thr Leu Trp Glu Gly Asn Pro Gly Arg
   2440                2445                2450

TTT TGG AAC ACT ACC ATT GCG GTG TCA ATG GCT AAC ATT TTT AGA GGG     7506
Phe Trp Asn Thr Thr Ile Ala Val Ser Met Ala Asn Ile Phe Arg Gly
2455                2460                2465                2470

AGT TAC TTG GCC GGA GCT GGA CTT CTC TTT TCT ATT ATG AAG AAC ACA     7554
Ser Tyr Leu Ala Gly Ala Gly Leu Leu Phe Ser Ile Met Lys Asn Thr
                2475                2480                2485

ACC AAC ACA AGA AGG GGA ACT GGC AAC ATA GGA GAG ACG CTT GGA GAG     7602
Thr Asn Thr Arg Arg Gly Thr Gly Asn Ile Gly Glu Thr Leu Gly Glu
           2490                2495                2500

AAA TGG AAA AGC CGA TTG AAC GCA TTG GGA AAA AGT GAA TTC CAG ATC     7650
Lys Trp Lys Ser Arg Leu Asn Ala Leu Gly Lys Ser Glu Phe Gln Ile
      2505                2510                2515

TAC AAG AAA AGT GGA ATC CAG GAA GTG GAT AGA ACC TTA GCA AAA GAA     7698
Tyr Lys Lys Ser Gly Ile Gln Glu Val Asp Arg Thr Leu Ala Lys Glu
   2520                2525                2530

GGC ATT AAA AGA GGA GAA ACG GAC CAT CAC GCT GTG TCG CGA GGC TCA     7746
Gly Ile Lys Arg Gly Glu Thr Asp His His Ala Val Ser Arg Gly Ser
2535                2540                2545                2550

GCA AAA CTG AGA TGG TTC GTT GAG AGA AAC ATG GTC ACA CCA GAA GGG     7794
Ala Lys Leu Arg Trp Phe Val Glu Arg Asn Met Val Thr Pro Glu Gly
                2555                2560                2565

AAA GTA GTG GAC CTC GGT TGT GGC AGA GGA GGC TGG TCA TAC TAT TGT     7842
Lys Val Val Asp Leu Gly Cys Gly Arg Gly Gly Trp Ser Tyr Tyr Cys
           2570                2575                2580

GGA GGA CTA AAG AAT GTA AGA GAA GTC AAA GGC CTA ACA AAA GGA GGA     7890
Gly Gly Leu Lys Asn Val Arg Glu Val Lys Gly Leu Thr Lys Gly Gly
      2585                2590                2595
```

```
CCA GGA CAC GAA GAA CCC ATC CCC ATG TCA ACA TAT GGG TGG AAT CTA     7938
Pro Gly His Glu Glu Pro Ile Pro Met Ser Thr Tyr Gly Trp Asn Leu
   2600                2605                2610

GTG CGT CTT CAA AGT GGA GTT GAC GTT TTC TTC ATC CCG CCA GAA AAG     7986
Val Arg Leu Gln Ser Gly Val Asp Val Phe Phe Ile Pro Pro Glu Lys
2615                2620                2625                2630

TGT GAC ACA TTA TTG TGT GAC ATA GGG GAG TCA TCA CCA AAT CCC ACA     8034
Cys Asp Thr Leu Leu Cys Asp Ile Gly Glu Ser Ser Pro Asn Pro Thr
                2635                2640                2645

GTG GAA GCA GGA CGA ACA CTC AGA GTC CTT AAC TTA GTA GAA AAT TGG     8082
Val Glu Ala Gly Arg Thr Leu Arg Val Leu Asn Leu Val Glu Asn Trp
           2650                2655                2660

TTG AAC AAC AAC ACT CAA TTT TGC ATA AAG GTT CTC AAC CCA TAT ATG     8130
Leu Asn Asn Asn Thr Gln Phe Cys Ile Lys Val Leu Asn Pro Tyr Met
      2665                2670                2675

CCC TCA GTC ATA GAA AAA ATG GAA GCA CTA CAA AGG AAA TAT GGA GGA     8178
Pro Ser Val Ile Glu Lys Met Glu Ala Leu Gln Arg Lys Tyr Gly Gly
   2680                2685                2690

GCC TTA GTG AGG AAT CCA CTC TCA CGA AAC TCC ACA CAT GAG ATG TAC     8226
Ala Leu Val Arg Asn Pro Leu Ser Arg Asn Ser Thr His Glu Met Tyr
2695                2700                2705                2710

TGG GTA TCC AAT GCT TCC GGG AAC ATA GTG TCA TCA GTG AAC ATG ATT     8274
Trp Val Ser Asn Ala Ser Gly Asn Ile Val Ser Ser Val Asn Met Ile
                2715                2720                2725

TCA AGG ATG TTG ATC AAC AGA TTT ACA ATG AGA TAC AAG AAA GCC ACT     8322
Ser Arg Met Leu Ile Asn Arg Phe Thr Met Arg Tyr Lys Lys Ala Thr
           2730                2735                2740

TAC GAG CCG GAT GTT GAC CTC GGA AGC GGA ACC CGT AAC ATC GGG ATT     8370
Tyr Glu Pro Asp Val Asp Leu Gly Ser Gly Thr Arg Asn Ile Gly Ile
      2745                2750                2755

GAA AGT GAG ATA CCA AAC CTA GAT ATA ATT GGG AAA AGA ATA GAA AAA     8418
Glu Ser Glu Ile Pro Asn Leu Asp Ile Ile Gly Lys Arg Ile Glu Lys
   2760                2765                2770

ATA AAG CAA GAG CAT GAA ACA TCA TGG CAC TAT GAC CAA GAC CAC CCA     8466
Ile Lys Gln Glu His Glu Thr Ser Trp His Tyr Asp Gln Asp His Pro
2775                2780                2785                2790

TAC AAA ACG TGG GCA TAC CAT GGT AGC TAT GAA ACA AAA CAG ACT GGA     8514
Tyr Lys Thr Trp Ala Tyr His Gly Ser Tyr Glu Thr Lys Gln Thr Gly
                2795                2800                2805

TCA GCA TCA TCC ATG GTC AAC GGA GTG GTC AGG CTG CTG ACA AAA CCT     8562
Ser Ala Ser Ser Met Val Asn Gly Val Val Arg Leu Leu Thr Lys Pro
           2810                2815                2820

TGG GAC GTT GTC CCC ATG GTG ACA CAG ATG GCA ATG ACA GAC ACG ACT     8610
Trp Asp Val Val Pro Met Val Thr Gln Met Ala Met Thr Asp Thr Thr
      2825                2830                2835
```

```
CCA TTT GGA CAA CAG CGC GTT TTT AAA GAG AAA GTG GAC ACG AGA ACC     8658
Pro Phe Gly Gln Gln Arg Val Phe Lys Glu Lys Val Asp Thr Arg Thr
   2840                2845                2850

CAA GAA CCG AAA GAA GGC ACG AAG AAA CTA ATG AAA ATA ACA GCA GAG     8706
Gln Glu Pro Lys Glu Gly Thr Lys Lys Leu Met Lys Ile Thr Ala Glu
2855                2860                2865                2870

TGG CTT TGG AAA GAA TTA GGG AAG AAA AAG ACA CCC AGG ATG TGC ACC     8754
Trp Leu Trp Lys Glu Leu Gly Lys Lys Lys Thr Pro Arg Met Cys Thr
                2875                2880                2885

AGA GAA GAA TTC ACA AGA AAG GTG AGA AGC AAT GCA GCC TTG GGG GCC     8802
Arg Glu Glu Phe Thr Arg Lys Val Arg Ser Asn Ala Ala Leu Gly Ala
            2890                2895                2900

ATA TTC ACT GAT GAG AAC AAG TGG AAG TCG GCA CGT GAG GCT GTT GAA     8850
Ile Phe Thr Asp Glu Asn Lys Trp Lys Ser Ala Arg Glu Ala Val Glu
        2905                2910                2915

GAT AGT AGG TTT TGG GAG CTG GTT GAC AAG GAA AGG AAT CTC CAT CTT     8898
Asp Ser Arg Phe Trp Glu Leu Val Asp Lys Glu Arg Asn Leu His Leu
   2920                2925                2930

GAA GGA AAG TGT GAA ACA TGT GTG TAC AAC ATG ATG GGA AAA AGA GAG     8946
Glu Gly Lys Cys Glu Thr Cys Val Tyr Asn Met Met Gly Lys Arg Glu
2935                2940                2945                2950

AAG AAG CTA GGG GAA TTC GGC AAG GCA AAA GGC AGC AGA GCC ATA TGG     8994
Lys Lys Leu Gly Glu Phe Gly Lys Ala Lys Gly Ser Arg Ala Ile Trp
                2955                2960                2965

TAC ATG TGG CTT GGA GCA CGC TTC TTA GAG TTT GAA GCC CTA GGA TTC     9042
Tyr Met Trp Leu Gly Ala Arg Phe Leu Glu Phe Glu Ala Leu Gly Phe
            2970                2975                2980

TTA AAT GAA GAT CAC TGG TTC TCC AGA GAG AAC TCC CTG AGT GGA GTG     9090
Leu Asn Glu Asp His Trp Phe Ser Arg Glu Asn Ser Leu Ser Gly Val
        2985                2990                2995

GAA GGA GAA GGG CTG CAC AAG CTA GGT TAC ATT CTA AGA GAC GTG AGC     9138
Glu Gly Glu Gly Leu His Lys Leu Gly Tyr Ile Leu Arg Asp Val Ser
   3000                3005                3010

AAG AAA GAG GGA GGA GCA ATG TAT GCC GAT GAC ACC GCA GGA TGG GAT     9186
Lys Lys Glu Gly Gly Ala Met Tyr Ala Asp Asp Thr Ala Gly Trp Asp
3015                3020                3025                3030

ACA AGA ATC ACA CTA GAA GAC KKA AAA AAT GAA GAA ATG GTA ACA AAC     9234
Thr Arg Ile Thr Leu Glu Asp Xaa Lys Asn Glu Glu Met Val Thr Asn
                3035                3040                3045

CAC ATG GAA GGA GAA CAC AAG AAA CTA GCC GAG GCC ATT TTC AAA CTA     9282
His Met Glu Gly Glu His Lys Lys Leu Ala Glu Ala Ile Phe Lys Leu
            3050                3055                3060

ACG TAC CAA AAC AAG GTG GTG CGT GTG CAA AGA CCA ACA CCA AGA GGC     9330
Thr Tyr Gln Asn Lys Val Val Arg Val Gln Arg Pro Thr Pro Arg Gly
        3065                3070                3075
```

```
ACA GTA ATG GAC ATC ATA TCG AGA AGA GAC CAA AGA GGT AGT GGA CAA    9378
Thr Val Met Asp Ile Ile Ser Arg Arg Asp Gln Arg Gly Ser Gly Gln
   3080                3085                3090

GTT GGC ACC TAT GGA CTC AAT ACT TTC ACC AAT ATG GAA GCC CAA CTA    9426
Val Gly Thr Tyr Gly Leu Asn Thr Phe Thr Asn Met Glu Ala Gln Leu
3095                3100                3105                3110

ATC AGA CAG ATG GAG GGA GAA GGA GTC TTT AAA AGC ATT CAG CAC CTA    9474
Ile Arg Gln Met Glu Gly Glu Gly Val Phe Lys Ser Ile Gln His Leu
                3115                3120                3125

ACA ATC ACA GAA GAA ATC GCT GTG CAA AAC TGG TTA GCA AGA GTG GGG    9522
Thr Ile Thr Glu Glu Ile Ala Val Gln Asn Trp Leu Ala Arg Val Gly
           3130                3135                3140

CGC GAA AGG TTA TCA AGA ATG GCC ATC AGT GGA GAT GAT TGT GTT GTG    9570
Arg Glu Arg Leu Ser Arg Met Ala Ile Ser Gly Asp Asp Cys Val Val
       3145                3150                3155

AAA CCT TTA GAT GAC AGG TTC GCA AGC GCT TTA ACA GCT CTA AAT GAC    9618
Lys Pro Leu Asp Asp Arg Phe Ala Ser Ala Leu Thr Ala Leu Asn Asp
   3160                3165                3170

ATG GGA AAG ATT AGG AAA GAC ATA CAA CAA TGG GAA CCT TCA AGA GGA    9666
Met Gly Lys Ile Arg Lys Asp Ile Gln Gln Trp Glu Pro Ser Arg Gly
3175                3180                3185                3190

TGG AAT GAT TGG ACA CAA GTG CCC TTC TGT TCA CAC CAT TTC CAT GAG    9714
Trp Asn Asp Trp Thr Gln Val Pro Phe Cys Ser His His Phe His Glu
                3195                3200                3205

TTA ATC ATG AAA GAC GGT CGC GTA CTC GTT GTT CCA TGT AGA AAC CAA    9762
Leu Ile Met Lys Asp Gly Arg Val Leu Val Val Pro Cys Arg Asn Gln
           3210                3215                3220

GAT GAA CTG ATT GGC AGA GCC CGA ATC TCC CAA GGA GCA GGG TGG TCT    9810
Asp Glu Leu Ile Gly Arg Ala Arg Ile Ser Gln Gly Ala Gly Trp Ser
       3225                3230                3235

TTG CGG GAG ACG GCC TGT TTG GGG AAG TCT TAC GCC CAA ATG TGG AGC    9858
Leu Arg Glu Thr Ala Cys Leu Gly Lys Ser Tyr Ala Gln Met Trp Ser
   3240                3245                3250

TTG ATG TAC TTC CAC AGA CGC GAC CTC AGG CTG GCG GCA AAT GCT ATT    9906
Leu Met Tyr Phe His Arg Arg Asp Leu Arg Leu Ala Ala Asn Ala Ile
3255                3260                3265                3270

TGC TCG GCA GTA CCA TCA CAT TGG GTT CCA ACA AGT CGA ACA ACC TGG    9954
Cys Ser Ala Val Pro Ser His Trp Val Pro Thr Ser Arg Thr Thr Trp
                3275                3280                3285

TCC ATA CAT GCT AAA CAT GAA TGG ATG ACA ACG GAA GAC ATG CTG ACA   10002
Ser Ile His Ala Lys His Glu Trp Met Thr Thr Glu Asp Met Leu Thr
           3290                3295                3300

GTC TGG AAC AGG GTG TGG ATT CAA GAA AAC CCA TGG ATG GAA GAC AAA   10050
Val Trp Asn Arg Val Trp Ile Gln Glu Asn Pro Trp Met Glu Asp Lys
       3305                3310                3315
```

```
ACT CCA GTG GAA TCA TGG GAG GAA ATC CCA TAC TTG GGG AAA AGA GAA   10098
Thr Pro Val Glu Ser Trp Glu Glu Ile Pro Tyr Leu Gly Lys Arg Glu
   3320                3325                3330

GAC CAA TGG TGC GGC TCA TTG ATT GGG TTA ACA AGC AGG GCC ACC TGG   10146
Asp Gln Trp Cys Gly Ser Leu Ile Gly Leu Thr Ser Arg Ala Thr Trp
3335                3340                3345                3350

GCA AAG AAC ATC CAA GCA GCA ATA AAT CAA GTT AGA TCC CTT ATA GGC   10194
Ala Lys Asn Ile Gln Ala Ala Ile Asn Gln Val Arg Ser Leu Ile Gly
               3355                3360                3365

AAT GAA GAA TAC ACA GAT TAC ATG CCA TCC ATG AAA AGA TTC AGA AGA   10242
Asn Glu Glu Tyr Thr Asp Tyr Met Pro Ser Met Lys Arg Phe Arg Arg
            3370                3375                3380

GAA GAG GAA GAA GCA GGA GTT CTG TGG TAGAAAGCAA AACTAACATG AAACAAGG 10297
Glu Glu Glu Glu Ala Gly Val Leu Trp
        3385                3390

CTAGAAGTCA GGTCGGATTA AGCCATAGTA CGGAAAAAAC TATGCTACCT GTGAGCCCCG  10357
TCCAAGGACG TTAAAAGAAG TCAGGCCATC ATAAATGCCA TAGCTTGAGT AAACTATGCA  10417
GCCTGTAGCT CCACCTGAGA AGGTGTAAAA AATCCGGGAG GCCACAAACC ATGGAAGCTG  10477
TACGCATGGC GTAGTGGACT AGCGGTTAGA GAGGACCCCT CCCTTACAAA TCGCAGCAAC  10537
AATGGGGGCC CAAGGCGAGA TGAAGCTGTA GTCTCGCTGG AAGGACTAGA GGTTAGAGGA  10597
GACCCCCCCG AAACAAAAAA CAGCATATTG ACGCTGGGAA AGACCAGAGA TCCTGCTGTC  10657
TCCTCAGCAT CATTCCAGGC ACAGAACGCC AGAAAATGGA ATGGTGCTGT TGAATCAACA  10717
GGTTCT                                                             10723
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 25 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCCAGTCACG ACGTTGTAAA ACGAC 25

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 27 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GGATGTGCTG CAAGGCGATT AAGTTGG 27

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TGAGCGGATA ACAATTTCAC ACAGG 25

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GGCTTTACAC TTTATGCTTC CGGCTCG 27

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 75 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCGGATATTG GAATTCTCTA GAAATTTAAT ACGACTCACT ATAAGTTGTT AGTCTACGTG 60
GACCGACAAA GACAG 75

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 77 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CCAGTGAATT CGAGCTCACG CGTAAATTTA ATACGACTCA CTATAAGTTG TTAGTCTACG 60
TGGACCGACA AAGACAG 77

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AGTTGTTAGT CTACGTGGAC CGAC 24

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GACAGATTCT TTGAGGGAGC TGAGCTCAAC GTAG 34

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TCAATATGCT GAAACGCGAG AGAAACCG 28

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGGATTGTTA GGAAACGAAG GAACGC 26

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CCACCAACAG CAGGGATACT GAAAAGATGG GG 32

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TGCAGATCTG CGTCTCCTAT TCAAG 25

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CGTGAACATG TGTACCCTCA TGGCC 25

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TTGCACCAAC AGTCAATGTC TTCAGG 26

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

ACCAGAAGAC ATAGATTGTT GGTGC 25

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GCACCAACAG TCTATGTCTT CTGGC 25

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

ATGTTTCCAG GCCCCTTCTG ATGAC 25

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GCAGCAATCC TGGCATACAC CATAG 25

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGTTGACATA GTCTTAGAAC ATGGAAG 27

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CTTCCATGTT CTAAGACTAT GTCAACC 27

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTCTTAGAAC ATGGAAGTTG TGTGACGACG ATGGC 35

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ACAACAGAAT CTCGCTGCCC AACAC 25

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCAAACACTC CATGGTAGAC AGAGG 25

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCTCTGTCTA CCATGGAGTG TTTGC 25

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CCACATCCAT TTCCCCATCC TCTGTCT 27

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GGAAAGGGAG GCATTGTGAC CTGTGCTATG 30

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGAAATCAAA ATAACACCAC AGAGTTCC 28

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CTGCAGCAAC ACCATCTCAT TGAAGTCGAG GCCC 34

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GACTTCAATG AGATGGTGCT GCTGC 25

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GCAGCAGCAC CATCTCATTG AAGTC 25

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

AAGCTTGGCT GGTGCACAGG CAATGGTT 28

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TGGTAACGGC AGGTCTAGGA ACCATTG 27

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GGACATCTCA AGTGCAGGCT GAG 23

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CTCAGCCTGC ACTTGAGATG TCC 23

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GAAGGAAATA GCAGAAACAC AACATGG 27

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CCCTTCATAT TGTACTCTGA TAACTATTGT TCC 33

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CCTCCATTCG GAGACAGCTA CATCATCATA GG 32

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CCTATGATGA TGTAGCTGTC TCCGAATGGA GG 32

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

ATGGCCATTT TAGGTGACAC AGCCTGGGA 29

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

TGTAAACACT CCTCCCAGGG ATCCAAA 27

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CTCATAGGAG TCATTATCAC ATGGATAGG 29

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GGGGATTCTG GTTGGAACTT ATATTGTTCT GTCC 34

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

TGATTCAATT CTGGTGTTAT TTGTTTCCAC 30

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

AAGGAATCAT GCAGGCAGGA AAACG 25

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

ACTTCCAGCG AGTTCCAAGC TC 22

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

AACAGAGCCG TCCATGCCGA TATGG 25

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

TCCATTGCTC CAAAGGGTGT GT 22

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

AGCTTGAGAT GGACTTTGAT TTCTG 25

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GGTCTGATTT CCATCCCGTA CC 22

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GTCCTTTAGA GACCTGGGAA GAG 23

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GTTTTCTCAA GAGTAGTCCA GCTGC 25

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

ATCAATTGGC AGTGACTATC ATGGC 25

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

TGTTAAGAGC AGTGGAGAAA CGGAC 25

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GATTGAGACC TTTGATCGTC AACGC 25

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

TGACAGGACC ATTAGTGGCT GGAGG 25

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CGTGCTCACT GGACGATCGG CCGATTTGGA ACTG 34

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GGGCTGCTTC CTGATATTTC TGCC 24

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CCTGTGGGAA GTGAAGAAAC AACGG 25

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GCTCCATCTT CCAGTTCAGC CTTTCCCATG 30

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CTCCGGCTCC AATCTGAGAG TATCC 25

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

CCTAATATCA TATGGAGGAG GCTGG 25

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GAAGGAGAAG AAGTCCAGGT ATTGG 25

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

CTGTCGACAA TTGGAGATCC TGACG 25

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GTGGAGCATA TGTGAGTGCT ATAGC 25

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

TCTGACTATG GCCGGAAGGT ATCTC 25

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

ACATTAATCT TGGCCCCCAC TAGAG 25

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CGATCTCCCG CCCGGTGTG 19

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

CTAACTGGTG ATAGCAGCCT CATGG 25

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

CCTACTGAGT TGTATCACTT TCTTTCC 27

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

TGGATTTCTT CCTATTCTCC CTCTTC 26

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

TTCAAGGCTG AGAGGGTTAT AGACC 25

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

TCTGGTTGGC CTACAGAGTG GCAGC 25

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

CCTTCTTTTG TCCAGATTTC CACTTCC 27

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

GCGTACAACC ATGCTCTCAG TGAACTGCCG GAGAC 35

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

TTCCCAGGGT CATCTTCCCT ATAC 24

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

GATGCTAGCC GTGATTATGC AGCACATTCC C 31

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

AAACAGAGAA CACCCCAAGA CAACC 25

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

CGGCATACAG CGTCCATGCT G 21

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

GTCTCGGGAA AGGATGGCCA TTGTC 25

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

CTCTGGTTGC TTTTGCTTGA AGTCC 25

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

CCGCCGCTGC TCTTTTCTGA GCTTCTC 27

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

AGGACTACAT GGGCTCTGTG TGAGG 25

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

GAGAAGTCCA GCTCCGGCC 19

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

AGAGAAACAT GGTCACACCA GAAGG 25

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

GTTCTTCGTG TCCTGGTCCT CC 22

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

GGAAATATGG AGGAGCCTAG TGAGG 25

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

ACCCAGTACA TCTCATGTGT GG 22

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

GAGCATGAAA CATCATGGCA CTATGACC 28

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

TCATGGCACT ATGACCAAGA CCACC 25

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

CAGTCTGACC ACTCCGTTCA CC 22

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

AAGGTGAGAA GCAATGCAGC CTTGG 25

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

GGGCCATATT CACTGATGAG AACAAGTGG 29

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

TCTTTCCCTG TCAACCAGCT CC 22

(2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

AATGAAGATC ACTGGTTCTC CAGAG 25

(2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

ACGTGAGCAA GAAAGAGGGA GGAGC 25

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

TGTCCCATCC TGCTGTGTCA TC 22

(2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

GCTAGTTTCT TGTGTTCTCC TTCCATGTGG 30

(2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

TCATATCGAG AAGAGACCAA AGAGG 25

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

ACTCCTTCTC CCTCCATCTG TCTG 24

(2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

ATGCTTTTGA AGATTCCTTC TCCCTCC 27

(2) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

GCACAGCGAT TTCTTCTGTG ATTGTTAGGT GC 32

(2) INFORMATION FOR SEQ ID NO:104:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

ACAATGGGAA CCTTCAAGAG GATGG 25

(2) INFORMATION FOR SEQ ID NO:105:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

TTATCACATT GGATCCTTCA AGAGGATGGA ATGATTGGAC ACAAG 45

(2) INFORMATION FOR SEQ ID NO:106:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

CAGAAGGGCA CTTGTGTCCA ATCATTCC 28

(2) INFORMATION FOR SEQ ID NO:107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

CTCCCTGGGA AATTCGGGCT C 21

(2) INFORMATION FOR SEQ ID NO:108:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

CCGTCTCCCG CAAAGACCAC CCTGCTCC 28

(2) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

TTATCACCTA TCTAGACCGT CTCCCGCAAA GACCACCCTG CTCC 44

(2) INFORMATION FOR SEQ ID NO:110:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

GTTGGAACCC AATGTGATGG TACTGC 26

(2) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

ACAAGTCGAA CAACCTGGTC CATAC 25

(2) INFORMATION FOR SEQ ID NO:112:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

GCATGTCTTC CGTCGTCATC C 21

(2) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

CTTGAATCCA CACCCTGTTC CAGAC 25

(2) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

ATACACAGAT TACATGCCAT CCATG 25

(2) INFORMATION FOR SEQ ID NO:115:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

TTTTGCCTTC TACCACAGGA C 21

(2) INFORMATION FOR SEQ ID NO:116:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

GAAACAAGGC TAGAAGTCAG GTCGG 25

(2) INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

GACGGGGCTC ACAGGTAGCA TAG 23

(2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

GCCTGTAGCT CCACCTGAGA AGGTG 25

(2) INFORMATION FOR SEQ ID NO:119:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

GGAAGCTGTA CGCATGGCGT AGTGG 25

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

GGGCCCCCGT TGTTGCTGC 19

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

AGAACCTGTT GATTCAACAG CACCATTCCA TTTTCTG 37

(2) INFORMATION FOR SEQ ID NO:122:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 59 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

TTATCACCTA GCATGCTCTA GAAGAACCTG TTGATTCAAC AGCACCATTC CATTTTCTG 59

(2) INFORMATION FOR SEQ ID NO:123:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

TTATCACCTA TCTAGAGAAC CTGTTGATTC AACAGCACCA TTCCATTTTC TG 52

(2) INFORMATION FOR SEQ ID NO:124:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2394 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...2394
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

```
AGA TTC TCA AAA GGA TTG CTC TCA GGC CAA GGA CCC ATG AAA TTG GTG      48
Arg Phe Ser Lys Gly Leu Leu Ser Gly Gln Gly Pro Met Lys Leu Val
 1               5                  10                  15
```

```
ATG GCT TTC ATA GCA TTC TTA AGA TTT CTA GCC ATA CCC CCA ACA GCA      96
Met Ala Phe Ile Ala Phe Leu Arg Phe Leu Ala Ile Pro Pro Thr Ala
            20                  25                  30

GGA ATT TTG GCT AGA TGG GGC TCA TTC AAG AAG AAT GGA GCG ATT AAA     144
Gly Ile Leu Ala Arg Trp Gly Ser Phe Lys Lys Asn Gly Ala Ile Lys
        35                  40                  45

GTG TTA CGG GGT TTC AAG AGA GAA ATC TCA AAC ATG CTA AAC ATA ATG     192
Val Leu Arg Gly Phe Lys Arg Glu Ile Ser Asn Met Leu Asn Ile Met
    50                  55                  60

AAC AGG AGG AAA AGA TCC GTG ACC ATG CTC CTT ATG CTG CTG CCC ACA     240
Asn Arg Arg Lys Arg Ser Val Thr Met Leu Leu Met Leu Leu Pro Thr
65                  70                  75                  80

GCC CTG GCG TTC CAT CTG ACG ACA CGA GGG GGA GAG CCG CAT ATG ATA     288
Ala Leu Ala Phe His Leu Thr Thr Arg Gly Gly Glu Pro His Met Ile
                85                  90                  95

GTT AGC AAG CAG GAA AGA GGA AAG TCA CTT TTG TTC AAG ACC TCT GCA     336
Val Ser Lys Gln Glu Arg Gly Lys Ser Leu Leu Phe Lys Thr Ser Ala
            100                 105                 110

GGT GTC AAC ATG TGC ACC CTC ATT GCG ATG GAT TTG GGA GAG TTG TGT     384
Gly Val Asn Met Cys Thr Leu Ile Ala Met Asp Leu Gly Glu Leu Cys
        115                 120                 125

GAG GAC ACG ATG ACC TAC AAA TGC CCC CGG ATC ACT GAG GCG GAA CCA     432
Glu Asp Thr Met Thr Tyr Lys Cys Pro Arg Ile Thr Glu Ala Glu Pro
    130                 135                 140

GAT GAC GTT GAC TGT TGG TGC AAT GCC ACG GAC ACA TGG GTG ACC TAT     480
Asp Asp Val Asp Cys Trp Cys Asn Ala Thr Asp Thr Trp Val Thr Tyr
145                 150                 155                 160

GGA ACG TGC TCT CAA ACT GGC GAA CAC CGA CGA GAC AAA CGT TCC GTC     528
Gly Thr Cys Ser Gln Thr Gly Glu His Arg Arg Asp Lys Arg Ser Val
                165                 170                 175

GCA TTG GCC CCA CAC GTG GGG CTT GGC CTA GAA ACA AGA GCC GAA ACG     576
Ala Leu Ala Pro His Val Gly Leu Gly Leu Glu Thr Arg Ala Glu Thr
            180                 185                 190

TGG ATG TCC TCT GAA GGT GCT TGG AAA CAG ATA CAA AAA GTA GAG ACT     624
Trp Met Ser Ser Glu Gly Ala Trp Lys Gln Ile Gln Lys Val Glu Thr
        195                 200                 205

TGG GCT CTG AGA CAT CCA GGA TTC ACG GTG ATA GCC CTT TTT CTA GCA     672
Trp Ala Leu Arg His Pro Gly Phe Thr Val Ile Ala Leu Phe Leu Ala
    210                 215                 220

CAT GCC ATA GGA ACA TCC ATC ACC CAG AAA GGG ATC ATT TTC ATT TTG     720
His Ala Ile Gly Thr Ser Ile Thr Gln Lys Gly Ile Ile Phe Ile Leu
225                 230                 235                 240

CTG ATG CTG GTA ACA CCA TCT ATG GCC ATG CGA TGC GTG GGA ATA GGC     768
Leu Met Leu Val Thr Pro Ser Met Ala Met Arg Cys Val Gly Ile Gly
                245                 250                 255
```

```
AAC AGA GAC TTC GTG GAA GGA CTG TCA GGA GCA ACA TGG GTG GAT GTG      816
Asn Arg Asp Phe Val Glu Gly Leu Ser Gly Ala Thr Trp Val Asp Val
            260                 265                 270

GTA CTG GAG CAT GGA AGT TGC GTC ACC ACC ATG GCA AAA AAC AAA CCA      864
Val Leu Glu His Gly Ser Cys Val Thr Thr Met Ala Lys Asn Lys Pro
        275                 280                 285

ACA CTG GAC ATT GAA CTC TTG AAG ACG GAG GTC ACA AAC CCT GCA GTT      912
Thr Leu Asp Ile Glu Leu Leu Lys Thr Glu Val Thr Asn Pro Ala Val
    290                 295                 300

CTG CGT AAA TTG TGC ATT GAA GCT AAA ATA TCA AAC ACC ACC ACC GAT      960
Leu Arg Lys Leu Cys Ile Glu Ala Lys Ile Ser Asn Thr Thr Thr Asp
305                 310                 315                 320

TCG AGA TGT CCA ACA CAA GGA GAA GCC ACA CTG GTG GAA GAA CAA GAC     1008
Ser Arg Cys Pro Thr Gln Gly Glu Ala Thr Leu Val Glu Glu Gln Asp
                325                 330                 335

GCG AAC TTT GTG TGC CGA CGA ACG TTC GTG GAC AGA GGC TGG GGC AAT     1056
Ala Asn Phe Val Cys Arg Arg Thr Phe Val Asp Arg Gly Trp Gly Asn
            340                 345                 350

GGC TGT GGG CTA TTC GGA AAA GGT AGT CTA ATA ACG TGT GCC AAG TTT     1104
Gly Cys Gly Leu Phe Gly Lys Gly Ser Leu Ile Thr Cys Ala Lys Phe
        355                 360                 365

AAG TGT GTG ACA AAA CTA GAA GGA AAG ATA GCT CAA TAT GAA AAC CTA     1152
Lys Cys Val Thr Lys Leu Glu Gly Lys Ile Ala Gln Tyr Glu Asn Leu
    370                 375                 380

AAA TAT TCA GTG ATA GTC ACC GTC CAC ACT GGA GAT CAG CAC CAG GTG     1200
Lys Tyr Ser Val Ile Val Thr Val His Thr Gly Asp Gln His Gln Val
385                 390                 395                 400

GGA AAT GAG ACT ACA GAA CAT GGA ACA ACT GCA ACC ATA ACA CCT CAA     1248
Gly Asn Glu Thr Thr Glu His Gly Thr Thr Ala Thr Ile Thr Pro Gln
                405                 410                 415

GCT CCT ACG TCG GAA ATA CAG CTG ACC GAC TAC GGA ACC CTT ACA TTA     1296
Ala Pro Thr Ser Glu Ile Gln Leu Thr Asp Tyr Gly Thr Leu Thr Leu
            420                 425                 430

GAT TGT TCA CCT AGG ACA GGG CTA GAT TTT AAC GAG ATG GTG TTG CTG     1344
Asp Cys Ser Pro Arg Thr Gly Leu Asp Phe Asn Glu Met Val Leu Leu
        435                 440                 445

ACA ATG AAA AAG AAA TCA TGG CTT GTC CAC AAA CAG TGG TTT CTA GAC     1392
Thr Met Lys Lys Lys Ser Trp Leu Val His Lys Gln Trp Phe Leu Asp
    450                 455                 460

TTA CCA CTG CCT TGG ACC TCT GGG GCT TTA ACA TCC CAA GAG ACT TGG     1440
Leu Pro Leu Pro Trp Thr Ser Gly Ala Leu Thr Ser Gln Glu Thr Trp
465                 470                 475                 480

AAC AGA CAA GAT TTA CTG GTC ACA TTT AAG ACA GCT CAT GCA AAG AAG     1488
Asn Arg Gln Asp Leu Leu Val Thr Phe Lys Thr Ala His Ala Lys Lys
                485                 490                 495
```

```
CAG GAA GTA GTC GTA CTA GGA TCA CAA GAA GGA GCA ATG CAC ACT GCG     1536
Gln Glu Val Val Val Leu Gly Ser Gln Glu Gly Ala Met His Thr Ala
            500                 505                 510

CTG ACT GGA GCG ACA GAA ATC CAA ACG TCA GGA ACG ACA ACA ATT TTC     1584
Leu Thr Gly Ala Thr Glu Ile Gln Thr Ser Gly Thr Thr Thr Ile Phe
        515                 520                 525

GCA GGA CAC CTA AAA TGC AGA CTA AAA ATG GAC AAA CTA ACT TTA AAA     1632
Ala Gly His Leu Lys Cys Arg Leu Lys Met Asp Lys Leu Thr Leu Lys
    530                 535                 540

GGG ATG TCA TAT GTG ATG TGC ACA GGC TCA TTC AAG TTA GAG AAA GAA     1680
Gly Met Ser Tyr Val Met Cys Thr Gly Ser Phe Lys Leu Glu Lys Glu
545                 550                 555                 560

GTG GCT GAG ACC CAG CAT GGA ACT GTT CTG GTG CAG GTT AAA TAT GAA     1728
Val Ala Glu Thr Gln His Gly Thr Val Leu Val Gln Val Lys Tyr Glu
                565                 570                 575

GGA ACA GAC GCA CCA TGC AAG ATT CCC TTT TCG ACC CAA GAT GAG AAA     1776
Gly Thr Asp Ala Pro Cys Lys Ile Pro Phe Ser Thr Gln Asp Glu Lys
            580                 585                 590

GGA GCA ACC CAG AAT GGG AGA TTA ATA ACA GCC AAC CCC ATA GTC ACT     1824
Gly Ala Thr Gln Asn Gly Arg Leu Ile Thr Ala Asn Pro Ile Val Thr
        595                 600                 605

GAC AAA GAA AAA CCA GTC AAT ATT GAG GCA GAA CCA CCC TTT GGT GAG     1872
Asp Lys Glu Lys Pro Val Asn Ile Glu Ala Glu Pro Pro Phe Gly Glu
    610                 615                 620

AGC TAC ATC GTG GTA GGA GCA GGT GAA AAA GCT TTG AAA CTA AGC TGG     1920
Ser Tyr Ile Val Val Gly Ala Gly Glu Lys Ala Leu Lys Leu Ser Trp
625                 630                 635                 640

TTC AAG AAA GGA AGC AGC ATA GGG AAA ATG TTT GAA GCA ACT GCC CGA     1968
Phe Lys Lys Gly Ser Ser Ile Gly Lys Met Phe Glu Ala Thr Ala Arg
                645                 650                 655

GGA GCA CGA AGG ATG GCC ATT CTG GGA GAC ACC GCA TGG GAC TTC GGT     2016
Gly Ala Arg Arg Met Ala Ile Leu Gly Asp Thr Ala Trp Asp Phe Gly
            660                 665                 670

TCT ATA GGA GGA GTG TTC ACG TCT ATG GGA AAA CTG GTA CAC CAG GTT     2064
Ser Ile Gly Gly Val Phe Thr Ser Met Gly Lys Leu Val His Gln Val
        675                 680                 685

TTT GGA ACT GCA TAT GGA GTT TTG TTT AGC GGA GTT TCT TGG ACC ATG     2112
Phe Gly Thr Ala Tyr Gly Val Leu Phe Ser Gly Val Ser Trp Thr Met
    690                 695                 700

AAA ATA GGA ATA GGG ATT CTG CTG ACA TGG CTA GGA TTA AAT TCA AGG     2160
Lys Ile Gly Ile Gly Ile Leu Leu Thr Trp Leu Gly Leu Asn Ser Arg
705                 710                 715                 720

AAC ACG TCC CTT TCG GTG ATG TGC ATC GCA GTT GGC ATG GTC ACA CTG     2208
Asn Thr Ser Leu Ser Val Met Cys Ile Ala Val Gly Met Val Thr Leu
                725                 730                 735
```

```
TAC CTA GGA GTC ATG GTT CAG GCA GAT TCG GGA TGT GTA ATC AAC TGG     2256
Tyr Leu Gly Val Met Val Gln Ala Asp Ser Gly Cys Val Ile Asn Trp
            740                 745                 750

AAA GGC AGA GAA CTT AAA TGT GGA AGC GGC ATT TTT GTC ACT AAT GAA     2304
Lys Gly Arg Glu Leu Lys Cys Gly Ser Gly Ile Phe Val Thr Asn Glu
        755                 760                 765

GTT CAC ACT TGG ACA GAG CAA TAC AAA TTC CAG GCT GAC TCC CCC AAG     2352
Val His Thr Trp Thr Glu Gln Tyr Lys Phe Gln Ala Asp Ser Pro Lys
    770                 775                 780

AGA CTA TCA GCA GCC ATT GGG AAG GCA TGG GAG GAG GGT GTG              2394
Arg Leu Ser Ala Ala Ile Gly Lys Ala Trp Glu Glu Gly Val
785                 790                 795
```

(2) INFORMATION FOR SEQ ID NO:125:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2145 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...2145
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

```
AAG GTC TTA AAA GGC TTC AAG AAG GAG ATC TCA AAC ATG CTG AGC ATT      48
Lys Val Leu Lys Gly Phe Lys Lys Glu Ile Ser Asn Met Leu Ser Ile
 1               5                  10                  15

ATC AAC AAA CGG AAA AAG ACA TCG CTC TGT CTC ATG ATG ATG TTA CCA      96
Ile Asn Lys Arg Lys Lys Thr Ser Leu Cys Leu Met Met Met Leu Pro
            20                  25                  30

GCA ACA CTT GCT TTC CAC TTA ACT TCA CGA GAT GGA GAG CCG CGC ATG     144
Ala Thr Leu Ala Phe His Leu Thr Ser Arg Asp Gly Glu Pro Arg Met
        35                  40                  45

ATT GTG GGG AAG AAT GAA AGA GGA AAA TCC CTA CTT TTC AAG ACA GCC     192
Ile Val Gly Lys Asn Glu Arg Gly Lys Ser Leu Leu Phe Lys Thr Ala
    50                  55                  60

TCT GGA ATC AAC ATG TGC ACA CTC ATA GCT ATG GAT CTG GGA GAG ATG     240
Ser Gly Ile Asn Met Cys Thr Leu Ile Ala Met Asp Leu Gly Glu Met
65                  70                  75                  80
```

```
TGT GAT GAC ACG GTC ACT TAC AAA TGC CCC CAC ATT ACC GAA GTG GAG      288
Cys Asp Asp Thr Val Thr Tyr Lys Cys Pro His Ile Thr Glu Val Glu
                85                  90                  95

CCT GAA GAC ATT GAC TGC TGG TGC AAC CTT ACA TCG ACA TGG GTG ACT      336
Pro Glu Asp Ile Asp Cys Trp Cys Asn Leu Thr Ser Thr Trp Val Thr
            100                 105                 110

TAT GGA ACA TGC AAT CAA GCT GGA GAG CAT AGA CGC GAT AAG AGA TCA      384
Tyr Gly Thr Cys Asn Gln Ala Gly Glu His Arg Arg Asp Lys Arg Ser
        115                 120                 125

GTG GCG TTA GCT CCC CAT GTT GGC ATG GGA CTG GAC ACA CGC ACT CAA      432
Val Ala Leu Ala Pro His Val Gly Met Gly Leu Asp Thr Arg Thr Gln
    130                 135                 140

ACC TGG ATG TCG GCT GAA GGA GCT TGG AGA CAA GTC GAG AAG GTA GAG      480
Thr Trp Met Ser Ala Glu Gly Ala Trp Arg Gln Val Glu Lys Val Glu
145                 150                 155                 160

ACA TGG GCC CTT AGG CAC CCA GGG TTT ACC ATA CTA GCC CTA TTT CTT      528
Thr Trp Ala Leu Arg His Pro Gly Phe Thr Ile Leu Ala Leu Phe Leu
                165                 170                 175

GCC CAT TAC ATA GGC ACT TCC TTG ACC CAG AAA GTG GTT ATT TTT ATA      576
Ala His Tyr Ile Gly Thr Ser Leu Thr Gln Lys Val Val Ile Phe Ile
            180                 185                 190

CTA TTA ATG CTG GTT ACC CCA TCC ATG ACA ATG AGA TGT GTA GGA GTA      624
Leu Leu Met Leu Val Thr Pro Ser Met Thr Met Arg Cys Val Gly Val
        195                 200                 205

GGA AAC AGA GAT TTT GTG GAA GGC CTA TCG GGA GCT ACG TGG GTT GAC      672
Gly Asn Arg Asp Phe Val Glu Gly Leu Ser Gly Ala Thr Trp Val Asp
    210                 215                 220

GTG GTG CTC GAG CAC GGT GGG TGT GTG ACT ACC ATG GCT AAG AAC AAG      720
Val Val Leu Glu His Gly Gly Cys Val Thr Thr Met Ala Lys Asn Lys
225                 230                 235                 240

CCC ACG CTG GAC ATA GAG CTT CAG AAG ACC GAG GCC ACC CAA CTG GCG      768
Pro Thr Leu Asp Ile Glu Leu Gln Lys Thr Glu Ala Thr Gln Leu Ala
                245                 250                 255

ACC CTA AGG AAG CTA TGC ATT GAG GGA AAA ATT ACC AAC ATA ACA ACC      816
Thr Leu Arg Lys Leu Cys Ile Glu Gly Lys Ile Thr Asn Ile Thr Thr
            260                 265                 270

GAC TCA AGA TGT CCC ACC CAA GGG GAA GCG ATT TTA CCT GAG GAG CAG      864
Asp Ser Arg Cys Pro Thr Gln Gly Glu Ala Ile Leu Pro Glu Glu Gln
        275                 280                 285

GAC CAG AAC TAC GTG TGT AAG CAT ACA TAC GTG GAC AGA GGC TGG GGA      912
Asp Gln Asn Tyr Val Cys Lys His Thr Tyr Val Asp Arg Gly Trp Gly
    290                 295                 300

AAC GGT TGT GGT TTG TTT GGC AAG GGA AGC TTG GTG ACA TGC GCG AAA      960
Asn Gly Cys Gly Leu Phe Gly Lys Gly Ser Leu Val Thr Cys Ala Lys
305                 310                 315                 320
```

```
TTT CAA TGT TTA GAA TCA ATA GAG GGA AAA GTG GTG CAA CAT GAG AAC    1008
Phe Gln Cys Leu Glu Ser Ile Glu Gly Lys Val Val Gln His Glu Asn
                325                 330                 335

CTC AAA TAC ACC GTC ATC ATC ACA GTG CAC ACA GGA GAC CAA CAC CAG    1056
Leu Lys Tyr Thr Val Ile Ile Thr Val His Thr Gly Asp Gln His Gln
            340                 345                 350

GTG GGA AAT GAA ACG CAG GGA GTC ACG GCT GAG ATA ACA CCC CAG GCA    1104
Val Gly Asn Glu Thr Gln Gly Val Thr Ala Glu Ile Thr Pro Gln Ala
        355                 360                 365

TCA ACC GCT GAA GCC ATT TTA CCT GAA TAT GGA ACC CTC GGG CTA GAA    1152
Ser Thr Ala Glu Ala Ile Leu Pro Glu Tyr Gly Thr Leu Gly Leu Glu
    370                 375                 380

TGC TCA CCA CGG ACA GGT TTG GAT TTC AAT GAA ATG ATC TCA TTG ACA    1200
Cys Ser Pro Arg Thr Gly Leu Asp Phe Asn Glu Met Ile Ser Leu Thr
385                 390                 395                 400

ATG AAG AAC AAA GCA TGG ATG GTA CAT AGA CAA TGG TTC TTT GAC TTA    1248
Met Lys Asn Lys Ala Trp Met Val His Arg Gln Trp Phe Phe Asp Leu
                405                 410                 415

CCC CTA CCA TGG ACA TCA GGA GCT ACA GCA GAA ACA CCA ACT TGG AAC    1296
Pro Leu Pro Trp Thr Ser Gly Ala Thr Ala Glu Thr Pro Thr Trp Asn
            420                 425                 430

AGG AAA GAG CTT CTT GTG ACA TTT AAA AAT GCA CAT GCA AAA AAG CAA    1344
Arg Lys Glu Leu Leu Val Thr Phe Lys Asn Ala His Ala Lys Lys Gln
        435                 440                 445

GAA GTA GTT GTT CTT GGA TCA CAA GAG GGA GCA ATG CAT ACA GCA CTG    1392
Glu Val Val Val Leu Gly Ser Gln Glu Gly Ala Met His Thr Ala Leu
    450                 455                 460

ACA GGA GCT ACA GAG ATC CAA ACC TCA GGA GGC ACA AGT ATC TTT GCG    1440
Thr Gly Ala Thr Glu Ile Gln Thr Ser Gly Gly Thr Ser Ile Phe Ala
465                 470                 475                 480

GGG CAC TTA AAA TGT AGA CTC AAG ATG GAC AAA TTG GAA CTC AAA GGG    1488
Gly His Leu Lys Cys Arg Leu Lys Met Asp Lys Leu Glu Leu Lys Gly
                485                 490                 495

ATG AGC TAT GCA ATG TGC TTG GGT AGC TTT GTG TTG AAG AAA GAA GTC    1536
Met Ser Tyr Ala Met Cys Leu Gly Ser Phe Val Leu Lys Lys Glu Val
            500                 505                 510

TCC GAA ACG CAG CAT GGG ACA ATA CTC ATT AAG GTT GAG TAC AAA GGG    1584
Ser Glu Thr Gln His Gly Thr Ile Leu Ile Lys Val Glu Tyr Lys Gly
        515                 520                 525

AAA GAT GCA CCC TGC AAG ATT CCT TTC TCC ACG GAG GAT GGA CAA GGA    1632
Lys Asp Ala Pro Cys Lys Ile Pro Phe Ser Thr Glu Asp Gly Gln Gly
    530                 535                 540

AAA GCT CAC AAT GGC AGA CTG ATC ACA GCC AAT CCA GTG GTG ACC AAG    1680
Lys Ala His Asn Gly Arg Leu Ile Thr Ala Asn Pro Val Val Thr Lys
545                 550                 555                 560
```

```
AAG GAG GAG CCT GTC AAC ATT GAG GCT GAA CCT CCT TTT GGA GAA AGT    1728
Lys Glu Glu Pro Val Asn Ile Glu Ala Glu Pro Pro Phe Gly Glu Ser
                565                 570                 575

AAC ATA GTA ATT GGA ATT GGA GAC AAA GCC CTG AAA ATC AAC TGG TAC    1776
Asn Ile Val Ile Gly Ile Gly Asp Lys Ala Leu Lys Ile Asn Trp Tyr
            580                 585                 590

AAG AAG GGA AGC TCG ATT GGG AAG ATG TTC GAG GCT ACT GCC AGA GGT    1824
Lys Lys Gly Ser Ser Ile Gly Lys Met Phe Glu Ala Thr Ala Arg Gly
        595                 600                 605

GCA AGG CGC ATG GCC ATC TTG GGA GAC ACA GCC TGG GAC TTT GGA TCA    1872
Ala Arg Arg Met Ala Ile Leu Gly Asp Thr Ala Trp Asp Phe Gly Ser
    610                 615                 620

GTG GGT GGT GTT TTG AAT TCA TTA GGG AAA ATG GTC CAC CAA ATA TTT    1920
Val Gly Gly Val Leu Asn Ser Leu Gly Lys Met Val His Gln Ile Phe
625                 630                 635                 640

GGG AGT GCT TAC ACA GCC CTA TTT GGT GGA GTC TCC TGG ATG ATG AAA    1968
Gly Ser Ala Tyr Thr Ala Leu Phe Gly Gly Val Ser Trp Met Met Lys
                645                 650                 655

ATT GGA ATA GGT GTC CTC TTA ACC TGG ATA GGG TTG AAC TCA AAA AAT    2016
Ile Gly Ile Gly Val Leu Leu Thr Trp Ile Gly Leu Asn Ser Lys Asn
            660                 665                 670

ACT TCT ATG TCA TTT TCA TGC ATC GCG ATA GGA ATC ATT ACA CTC TAT    2064
Thr Ser Met Ser Phe Ser Cys Ile Ala Ile Gly Ile Ile Thr Leu Tyr
        675                 680                 685

CTG GGA GCC GTG GTG CAA GCT GAC ATG GGG TGT GTC ATA AAC TGG AAA    2112
Leu Gly Ala Val Val Gln Ala Asp Met Gly Cys Val Ile Asn Trp Lys
    690                 695                 700

GGC AAA GAA CTC AAA TGT GGA AGT GGA ATT TTC                        2145
Gly Lys Glu Leu Lys Cys Gly Ser Gly Ile Phe
705                 710                 715
```

(2) INFORMATION FOR SEQ ID NO:126:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2175 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...2175
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

```
ATT CTG AAG AGA TGG GGA CAG TTG AAG AAA AAT AAG GCC ATC AGG ATA      48
Ile Leu Lys Arg Trp Gly Gln Leu Lys Lys Asn Lys Ala Ile Arg Ile
 1               5                  10                  15

CTG ATT GGA TTC AGG AAG GAG ATA GGC CGC ATG CTG AAC ATC TTG AAC      96
Leu Ile Gly Phe Arg Lys Glu Ile Gly Arg Met Leu Asn Ile Leu Asn
            20                  25                  30

GGG AGA AAA AGG TCA ACG ATA ACA TTG CTG TGC TTG ATT CCC ACC GTA     144
Gly Arg Lys Arg Ser Thr Ile Thr Leu Leu Cys Leu Ile Pro Thr Val
        35                  40                  45

ATG GCG TTT CAC TTG TCA ACA AGA GAT GGC GAA CCC CTC ATG ATA GTG     192
Met Ala Phe His Leu Ser Thr Arg Asp Gly Glu Pro Leu Met Ile Val
    50                  55                  60

GCA AAA CAT GAA AGG GGG AGA CCT CTC TTG TTT AAG ACA ACA GAG GGG     240
Ala Lys His Glu Arg Gly Arg Pro Leu Leu Phe Lys Thr Thr Glu Gly
65                  70                  75                  80

ATC AAC AAA TGC ACT CTC ATT GCC ATG GAC TTG GGT GAA ATG TGT GAG     288
Ile Asn Lys Cys Thr Leu Ile Ala Met Asp Leu Gly Glu Met Cys Glu
                85                  90                  95


GAC ACT GTC ACG TAT AAA TGC CCC TTA CTG GTC AAT ACC GAA CCT GAA     336
Asp Thr Val Thr Tyr Lys Cys Pro Leu Leu Val Asn Thr Glu Pro Glu
            100                 105                 110

GAC ATT GAT TGC TGG TGC AAT CTC ACG TCT ACC TGG GTC ACA TAT GGG     384
Asp Ile Asp Cys Trp Cys Asn Leu Thr Ser Thr Trp Val Thr Tyr Gly
        115                 120                 125

ACA TAC ACC CAG AGC GGA GAA CGG AGA CGA GAG AAG CGC TCA GTA GCT     432
Thr Tyr Thr Gln Ser Gly Glu Arg Arg Arg Glu Lys Arg Ser Val Ala
    130                 135                 140

TTA ACA CCA CAT TCA GGA ATG GGA TTG GAA ACA AGA GCT GAG ACA TGG     480
Leu Thr Pro His Ser Gly Met Gly Leu Glu Thr Arg Ala Glu Thr Trp
145                 150                 155                 160

ATG TCA TCG GAA GGG GCT TGG AAG CAT GCT CAG AGA GTA GAG AGC TGG     528
Met Ser Ser Glu Gly Ala Trp Lys His Ala Gln Arg Val Glu Ser Trp
                165                 170                 175

ATA CTC AGA AAC CCA GGA TTC GCG CTC TTG GCA GGA TTT ATG GCT TAT     576
Ile Leu Arg Asn Pro Gly Phe Ala Leu Leu Ala Gly Phe Met Ala Tyr
            180                 185                 190

ATG ATT GGG CAA ACA GGA ATC CAG CGA ACT GTC TTC TTT GTC CTA ATG     624
Met Ile Gly Gln Thr Gly Ile Gln Arg Thr Val Phe Phe Val Leu Met
        195                 200                 205

ATG CTG GTC GCC CCA TCC TAC GGA ATG CGA TGC GTA GGA GTA GGA AAC     672
Met Leu Val Ala Pro Ser Tyr Gly Met Arg Cys Val Gly Val Gly Asn
    210                 215                 220
```

```
AGA GAC TTT GTG GAA GGA GTC TCA GGT GGA GCA TGG GTC GAT CTG GTG      720
Arg Asp Phe Val Glu Gly Val Ser Gly Gly Ala Trp Val Asp Leu Val
225                 230                 235                 240

CTA GAA CAT GGA GGA TGC GTC ACA ACC ATG GCC CAG GGA AAA CCA ACC      768
Leu Glu His Gly Gly Cys Val Thr Thr Met Ala Gln Gly Lys Pro Thr
                245                 250                 255

TTG GAT TTT GAA CTG ACT AAG ACA ACA GCC AAG GAA GTG GCT CTG TTA      816
Leu Asp Phe Glu Leu Thr Lys Thr Thr Ala Lys Glu Val Ala Leu Leu
            260                 265                 270

AGA ACC TAT TGC ATT GAA GCC TCA ATA TCA AAC ATA ACC ACG GCA ACA      864
Arg Thr Tyr Cys Ile Glu Ala Ser Ile Ser Asn Ile Thr Thr Ala Thr
        275                 280                 285

AGA TGT CCA ACG CAA GGA GAG CCT TAT CTA AAA GAG GAA CAA GAC CAA      912
Arg Cys Pro Thr Gln Gly Glu Pro Tyr Leu Lys Glu Glu Gln Asp Gln
    290                 295                 300

CAG TAC ATT TGC CGG AGA GAT GTG GTA GAC AGA GGG TGG GGC AAT GGC      960
Gln Tyr Ile Cys Arg Arg Asp Val Val Asp Arg Gly Trp Gly Asn Gly
305                 310                 315                 320

TGT GGC TTG TTT GGA AAA GGA GGA GTT GTG ACA TGT GCG AAG TTT TCA     1008
Cys Gly Leu Phe Gly Lys Gly Gly Val Val Thr Cys Ala Lys Phe Ser
                325                 330                 335

TGT TCG GGG AAG ATA ACA GGC AAT TTG GTC CAA ATT GAG AAC CTT GAA     1056
Cys Ser Gly Lys Ile Thr Gly Asn Leu Val Gln Ile Glu Asn Leu Glu
            340                 345                 350

TAC ACA GTG GTT GTA ACA GTC CAC AAT GGA GAC ACC CAT GCA GTA GGA     1104
Tyr Thr Val Val Val Thr Val His Asn Gly Asp Thr His Ala Val Gly
        355                 360                 365

AAT GAC ACA TCC AAT CAT GGA GTT ACA GCC ACG ATA ACT CCC AGG TCA     1152
Asn Asp Thr Ser Asn His Gly Val Thr Ala Thr Ile Thr Pro Arg Ser
    370                 375                 380

CCA TCG GTG GAA GTC AAA TTG CCG GAC TAT GGA GAA CTA ACA CTC GAT     1200
Pro Ser Val Glu Val Lys Leu Pro Asp Tyr Gly Glu Leu Thr Leu Asp
385                 390                 395                 400

TGT GAA CCC AGG TCT GGA ATT GAC TTT AAT GAG ATG ATT CTG ATG AAA     1248
Cys Glu Pro Arg Ser Gly Ile Asp Phe Asn Glu Met Ile Leu Met Lys
                405                 410                 415

ATG AAA AAG AAA ACA TGG CTT GTG CAT AAG CAA TGG TTT TTG GAT CTA     1296
Met Lys Lys Lys Thr Trp Leu Val His Lys Gln Trp Phe Leu Asp Leu
            420                 425                 430

CCT CTA CCA TGG ACA GCA GGA GCA GAC ACA TCA GAG GTT CAC TGG AAT     1344
Pro Leu Pro Trp Thr Ala Gly Ala Asp Thr Ser Glu Val His Trp Asn
        435                 440                 445

TAC AAA GAG AGA ATG GTG ACA TTT AAG GTT CCT CAT GCC AAG AGA CAG     1392
Tyr Lys Glu Arg Met Val Thr Phe Lys Val Pro His Ala Lys Arg Gln
    450                 455                 460
```

```
GAT GTG ACA GTG CTG GGA TCT CAG GAA GGA GCC ATG CAT TCT GCC CTC    1440
Asp Val Thr Val Leu Gly Ser Gln Glu Gly Ala Met His Ser Ala Leu
465                 470                 475                 480

GCT GGA GCC ACA GAA GTG GAC TCC GGT GAT GGA AAT CAC ATG TTT GCA    1488
Ala Gly Ala Thr Glu Val Asp Ser Gly Asp Gly Asn His Met Phe Ala
                485                 490                 495

GGA CAT CTC AAG TGC AAA GTC CGT ATG GAG AAA TTG AGA ATC AAG GGA    1536
Gly His Leu Lys Cys Lys Val Arg Met Glu Lys Leu Arg Ile Lys Gly
            500                 505                 510

ATG TCA TAC ACG ATG TGT TCA GGA AAG TTC TCA ATT GAC AAA GAG ATG    1584
Met Ser Tyr Thr Met Cys Ser Gly Lys Phe Ser Ile Asp Lys Glu Met
        515                 520                 525

GCA GAA ACA CAG CAT GGG ACA ACA GTG GTG AAA GTC AAG TAT GAA GGT    1632
Ala Glu Thr Gln His Gly Thr Thr Val Val Lys Val Lys Tyr Glu Gly
    530                 535                 540

GCT GGA GCT CCG TGT AAA GTC CCC ATA GAG ATA AGA GAT GTG AAC AAG    1680
Ala Gly Ala Pro Cys Lys Val Pro Ile Glu Ile Arg Asp Val Asn Lys
545                 550                 555                 560

AAA AAA GTG GTT GGG CGT ATC ATC TCA TCC ACC CCT TTG GCT GAG AAT    1728
Lys Lys Val Val Gly Arg Ile Ile Ser Ser Thr Pro Leu Ala Glu Asn
                565                 570                 575

ACC AAC AGT GCA ACC AAC ATA GAG TTA GAA CCC CCC TTT GGG GAC AGC    1776
Thr Asn Ser Ala Thr Asn Ile Glu Leu Glu Pro Pro Phe Gly Asp Ser
            580                 585                 590

TAC ATA GTG ATA GGT GTT GGA AAC AGT GCA TTA ACA CTC CAT TGG TTC    1824
Tyr Ile Val Ile Gly Val Gly Asn Ser Ala Leu Thr Leu His Trp Phe
        595                 600                 605

AGG AAA GGG AGT TCC ATT GGC AAG ATG TTT GAG TCC ACA TAC AGA GGT    1872
Arg Lys Gly Ser Ser Ile Gly Lys Met Phe Glu Ser Thr Tyr Arg Gly
    610                 615                 620

GCA AAA CGA ATG GCC ATT CTA GGT GAA ACA GCT TGG GAT TTT GGT TCC    1920
Ala Lys Arg Met Ala Ile Leu Gly Glu Thr Ala Trp Asp Phe Gly Ser
625                 630                 635                 640

GTT GGT GGA CTG TTC ACA TCA TTG GGA AAG GCT GTG CAC CAG GTT TTT    1968
Val Gly Gly Leu Phe Thr Ser Leu Gly Lys Ala Val His Gln Val Phe
                645                 650                 655

GGA AGT GTG TAT ACA ACC ATG TTT GGA GGA GTC TCA TGG ATG ATT AGA    2016
Gly Ser Val Tyr Thr Thr Met Phe Gly Gly Val Ser Trp Met Ile Arg
            660                 665                 670

ATC CTA ATT GGG TTC CTA GTG TTG TGG ATT GGC ACG AAC TCA AGG AAC    2064
Ile Leu Ile Gly Phe Leu Val Leu Trp Ile Gly Thr Asn Ser Arg Asn
        675                 680                 685

ACT TCA ATG GCT ATG ACG TGC ATA GCT GTT GGA GGA ATC ACT CTG TTT    2112
Thr Ser Met Ala Met Thr Cys Ile Ala Val Gly Gly Ile Thr Leu Phe
    690                 695                 700
```

```
CTG GGC TTC ACA GTT CAA GCA GAG ATG GGT TGT GTG GTG TCA TGG AGT    2160
Leu Gly Phe Thr Val Gln Ala Glu Met Gly Cys Val Val Ser Trp Ser
705                 710                 715                 720

GGG AAA GAA TTG AGG                                            2175
Gly Lys Glu Leu Arg
                725
```

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

CACTACCGCA AGGTAGAGAG CTCGGCATTG CCTCTTGGTG 40

(2) INFORMATION FOR SEQ ID NO:128:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

GTGATGGCGT TCCATCTCTC GAGCCGTAAC GGAGAACCAC 40

(2) INFORMATION FOR SEQ ID NO:129:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

GCCCTGGCGT TCCATCTCTC GAGCCGAGGG GGAGAGCCGC 40

(2) INFORMATION FOR SEQ ID NO:130:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

ACACTTGCTT TCCACCTCTC GAGCCGAGAT GGAGAGCCGC 40

(2) INFORMATION FOR SEQ ID NO:131:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

GTAATGGCGT TTCACCTCTC GAGCAGAGAT GGCGAACCCC 40

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

CCTATCCTTA CTTAAGATCT TCGTGGAGTG ACAGAC 36

(2) INFORMATION FOR SEQ ID NO:133:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

GGATAGGAAT GAATTCTAGA AGCACCTCAC TGTCTG 36

(2) INFORMATION FOR SEQ ID NO:134:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

CCGCAGAGAT CGTTTTCCTG CCTGCATGAT TCC 33

(2) INFORMATION FOR SEQ ID NO:135:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

CCGATCCTAA TTTAAGATCT TTGTGCAGGG AAAGCC 36

(2) INFORMATION FOR SEQ ID NO:136:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

CCTATCCCAA CTTGAGATCT TTATGAAGAT ACAGTA 36

(2) INFORMATION FOR SEQ ID NO:137:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

CCTAACCGTG CTTGAGATCT TTGTGAAGTT ACCGAC 36

WHAT IS CLAIMED IS:

1. A quadravalent vaccine providing immunity against all four serotypes of dengue virus comprising a DEN-2 PDK-53 infectious clone-derived virus.

2. A quadravalent vaccine providing immunity against all four serotypes of dengue virus comprising a chimeric DEN-2/1 virus.

3. A quadravalent vaccine providing immunity against all four serotypes of dengue virus comprising a chimeric DEN-2/3 virus.

4. A quadravalent vaccine providing immunity against all four serotypes of dengue virus comprising a chimeric DEN-2/4 virus.

5. A quadravalent vaccine providing immunity against all four serotypes of dengue virus comprising DEN-2 PDK-53 infectious clone-derived and chimeric DEN-2/1, DEN-2/3, and DEN-2/4 viruses.

6. A method of immunization in which a desired immune response is produced against all four serotypes of dengue virus comprising the step of administering to a subject a quadravalent vaccine comprising DEN-2 PDK-53 infectious clone-derived and chimeric DEN-2/1, DEN-2/3, and DEN-2/4 viruses.

7. A composition of matter comprising a full genome-length infectious cDNA clone for a DEN-2 virus, strain 16681.

8. A composition of matter comprising a full genome-length infectious cDNA clone for a DEN-2 virus of a

strain characterized as replicating to high titer in cell culture.

9. A composition of matter comprising a full genome-length infectious cDNA clone for a DEN-2 virus, strain 16681, having the identifying characteristics of ATCC 69826.

10. A composition of matter comprising a full genome-length infectious cDNA clone for a DEN-2 virus, strain 16681, attenuated derivative, PDK-53.

11. A composition of matter comprising a full genome-length infectious cDNA clone for a DEN-2 virus attenuated derivative, characterized as replicating to high titer in cell culture.

12. A composition of matter comprising a full genome-length infectious cDNA clone for a DEN-2 virus, strain 16681, attenuated derivative, PDK-53, having the identifying characteristics of ATCC 69825.

13. A composition of matter comprising a full genome-length infectious cDNA clone of a chimeric DEN-2/1 virus, wherein said virus is characterized as the expressing prM and E genes of a DEN-1 attenuated virus in the context of the nonstructural genes of the DEN-2 PDK-53 virus.

14. The composition of matter of Claim 13, wherein said DEN-1 attenuated virus is DEN-1 PDK-13.

15. A composition of matter comprising a full genome-length infectious cDNA clone of a chimeric DEN-2 virus, wherein said virus is characterized as expressing the antigenicity of a DEN-1 attenuated virus.

16. A composition of matter comprising a full genome-length infectious cDNA clone of a chimeric DEN-2/3 virus, wherein said virus is characterized as expressing the prM and E genes of a DEN-3 attenuated virus in the context of the nonstructural genes of the DEN-2 PDK-53 virus.

17. The composition of matter of Claim 16, wherein said DEN-3 attenuated virus is DEN-3 PGMK30/FRhL-3.

18. A composition of matter comprising a full genome-length infectious cDNA clone of a chimeric DEN-2 virus, wherein said virus is characterized as expressing the antigenicity of a DEN-3 attenuated virus.

19. A composition of matter comprising a full genome-length infectious cDNA clone of a chimeric DEN-2/4 virus, wherein said virus is characterized as expressing the prM and E genes of a DEN-4 attenuated virus in the context of the nonstructural genes of the DEN-2 PDK-53 virus.

20. The composition of matter of Claim 19, wherein said DEN-4 attenuated virus is DEN-4 PDK-48.

21. A composition of matter comprising a full genome-length infectious cDNA clone of a chimeric DEN-2 virus, wherein said virus is characterized as expressing the antigenicity of a DEN-4 attenuated virus.

22. A genetic construct comprising a DNA sequence operably encoding the polyprotein of DEN-2 virus, strain 16681.

23. The genetic construct of Claim 22, wherein said polyprotein is the polyprotein encoded by the nucleotide sequence of SEQ ID NO:1.

24. A genetic construct comprising a DNA sequence operably encoding at least one protein of DEN-2 virus, strain 16681.

25. The genetic construct of Claim 24, wherein said protein is a protein encoded by the nucleotide sequence of SEQ ID NO: 1.

26. A genetic construct comprising a DNA sequence operably encoding the polyprotein of DEN-2 virus, strain 16681, attenuated derivative, PDK-53.

27. The genetic construct of Claim 26, wherein said polyprotein is the polyprotein encoded by the nucleotide sequence of SEQ ID NO:2.

28. A genetic construct comprising a DNA sequence operably encoding at least one protein of DEN-2 virus, strain 16681, attenuated derivative, PDK-53.

29. The genetic construct of Claim 28, wherein said protein is a protein encoded by the nucleotide sequence of SEQ ID NO: 2.

30. A genetic construct comprising a DNA sequence operably encoding at least one structural protein of DEN-1 PDK-13.

31. The genetic construct of Claim 30, wherein said structural protein is a structural protein encoded by the nucleotide sequence of SEQ ID NO: 124.

32. A genetic construct comprising a DNA sequence operably encoding at least one structural protein of DEN-3 PGMK30/FRhL-3.

33. The genetic construct of Claim 32, wherein said structural protein is a structural protein encoded by the nucleotide sequence of SEQ ID NO: 125.

34. A genetic construct comprising a DNA sequence operably encoding at least one structural protein of DEN-4 PDK-48.

35. The genetic construct of Claim 34, wherein said structural protein is a structural protein encoded by the nucleotide sequence of SEQ ID NO: 126.

36. A host cell comprising the genetic construct of any of Claims 22-35.

Construction of DEN-2 Infectious cDNA Clone

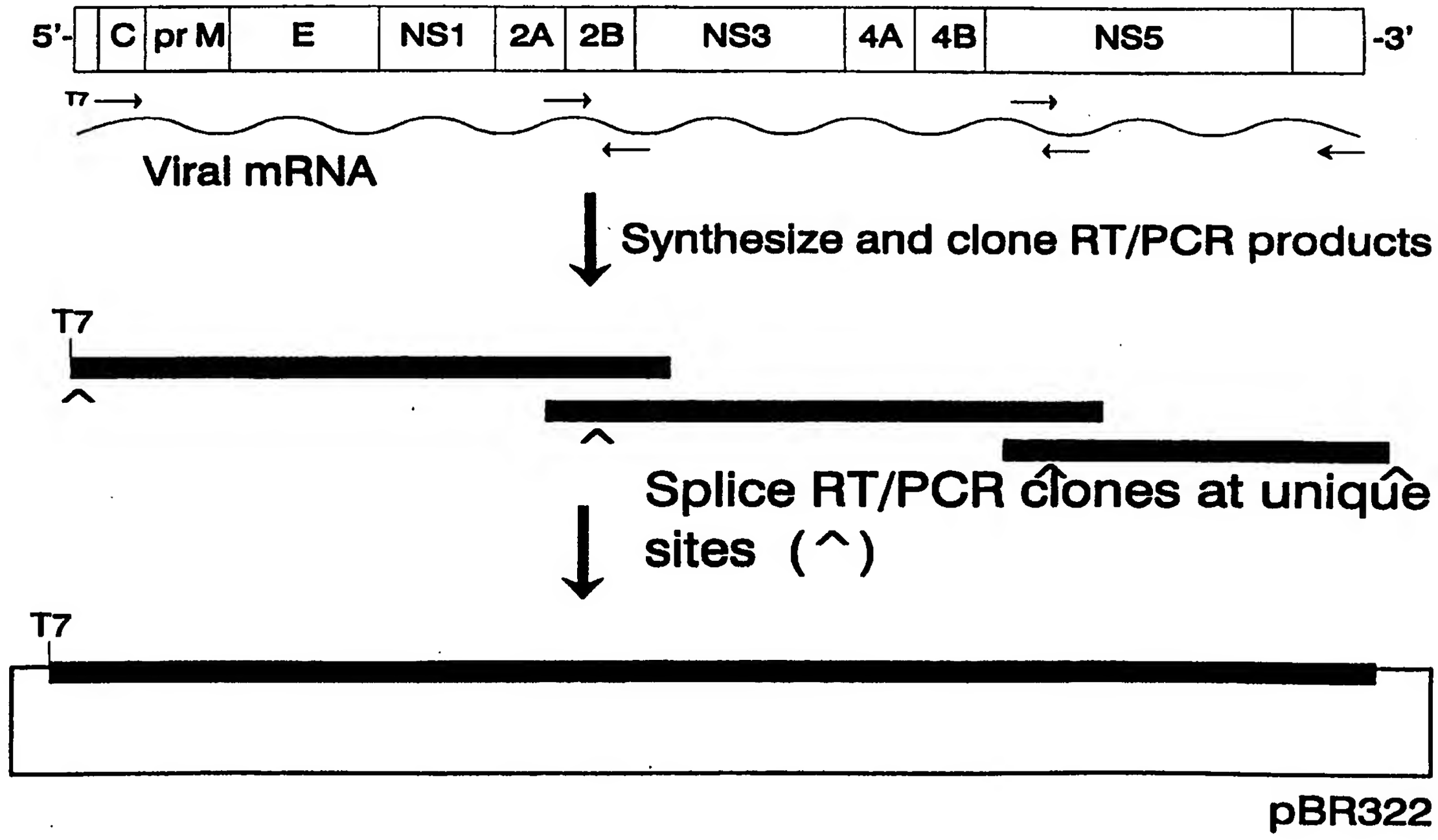


FIGURE 1

Transcription of DEN-2 RNA from Infectious Clone

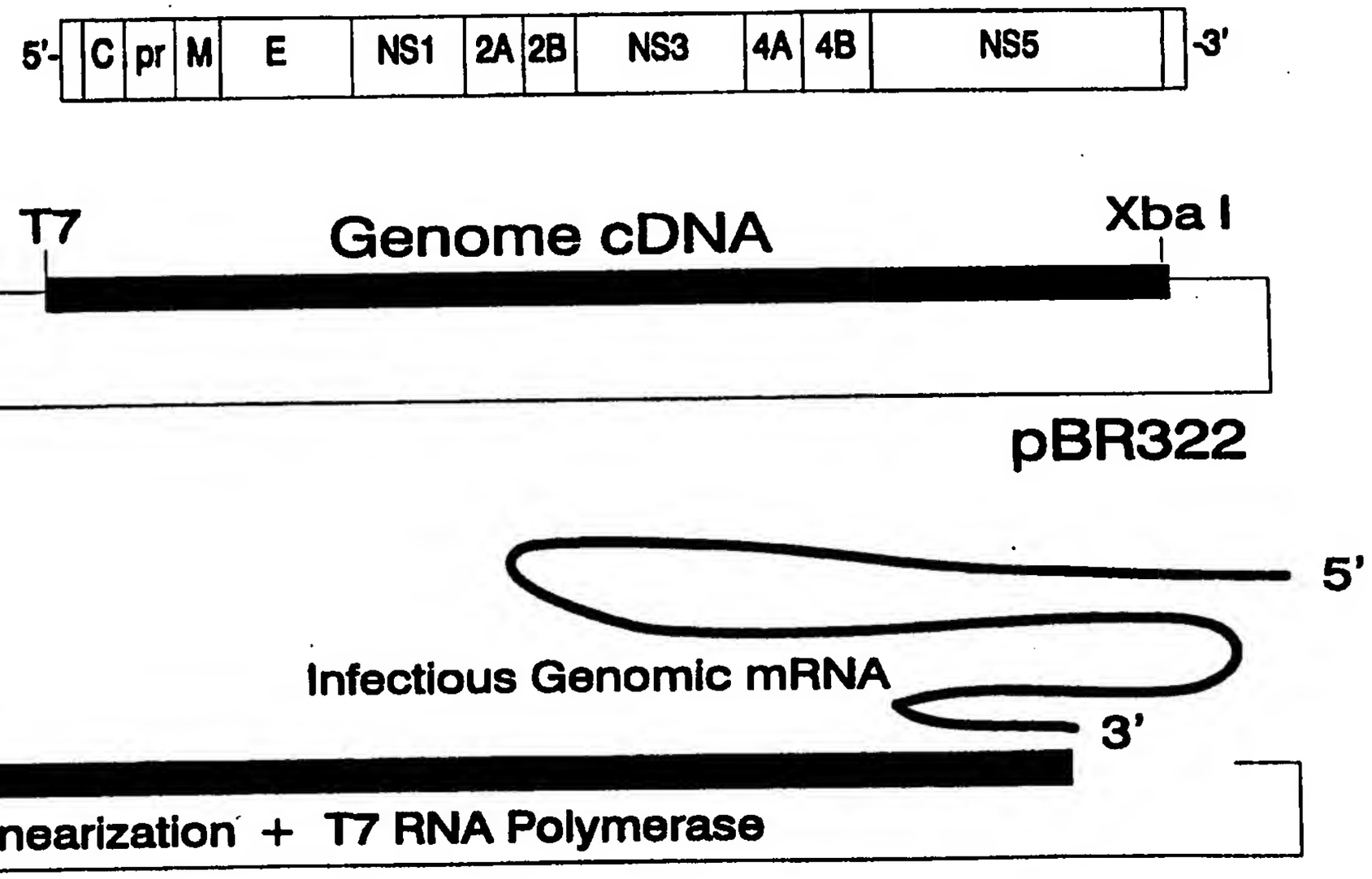


FIGURE 2

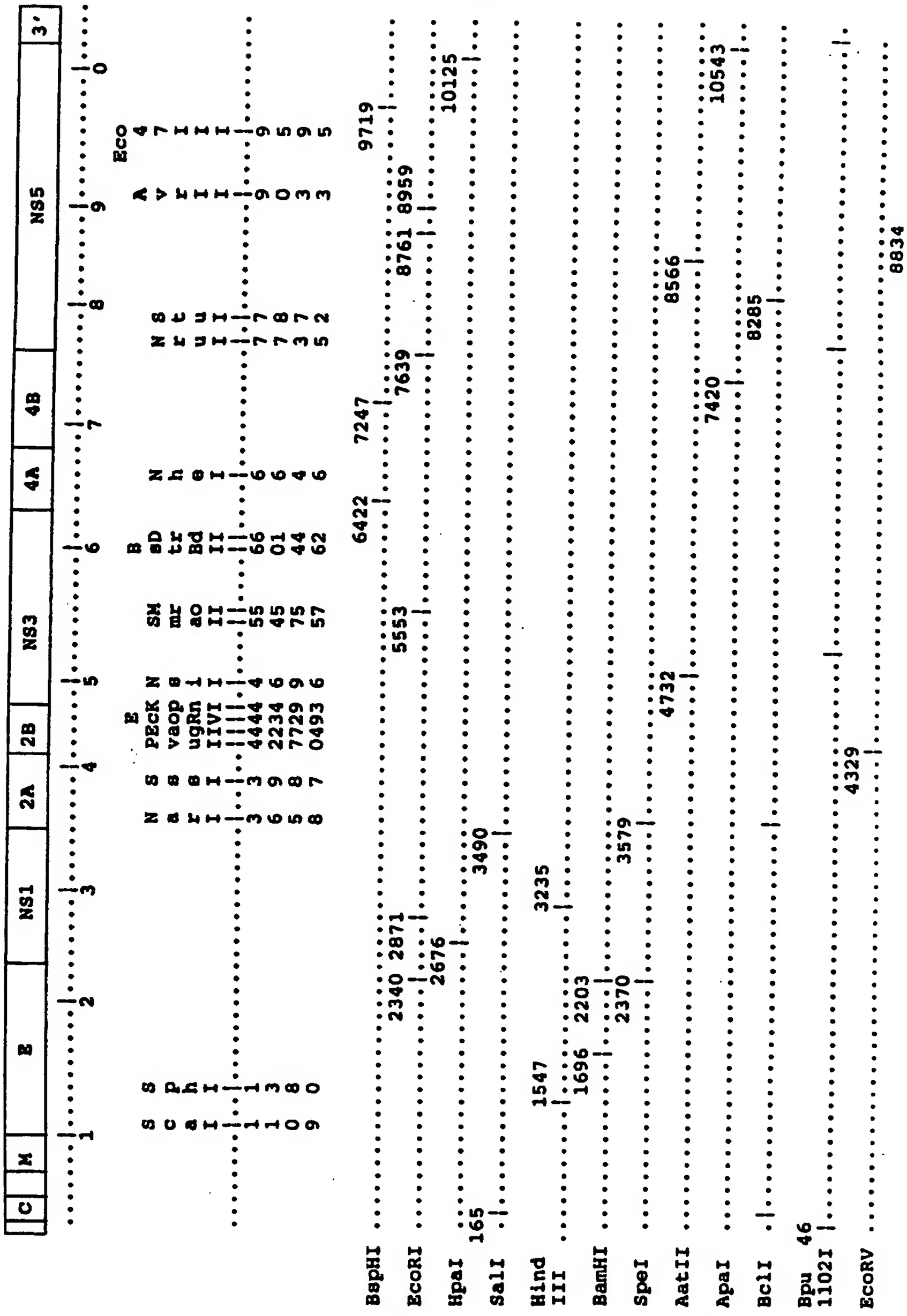

C
M
E
NS1
2A
2B
NS3
4A
4B
3'
BspHI
EcoRI
HpaI
SalI
HindIII
BamHI
SpeI
AatII
ApaI
BclI
Bpu1102I
EcoRV
9719
8761
8959
10125
8566
10543
8285
8834
6422
7247
5553
7639
7420
4732
4329
3579
3490
3235
2871
2676
2340
2203
2370
1547
1696
165
46

FIGURE 3A

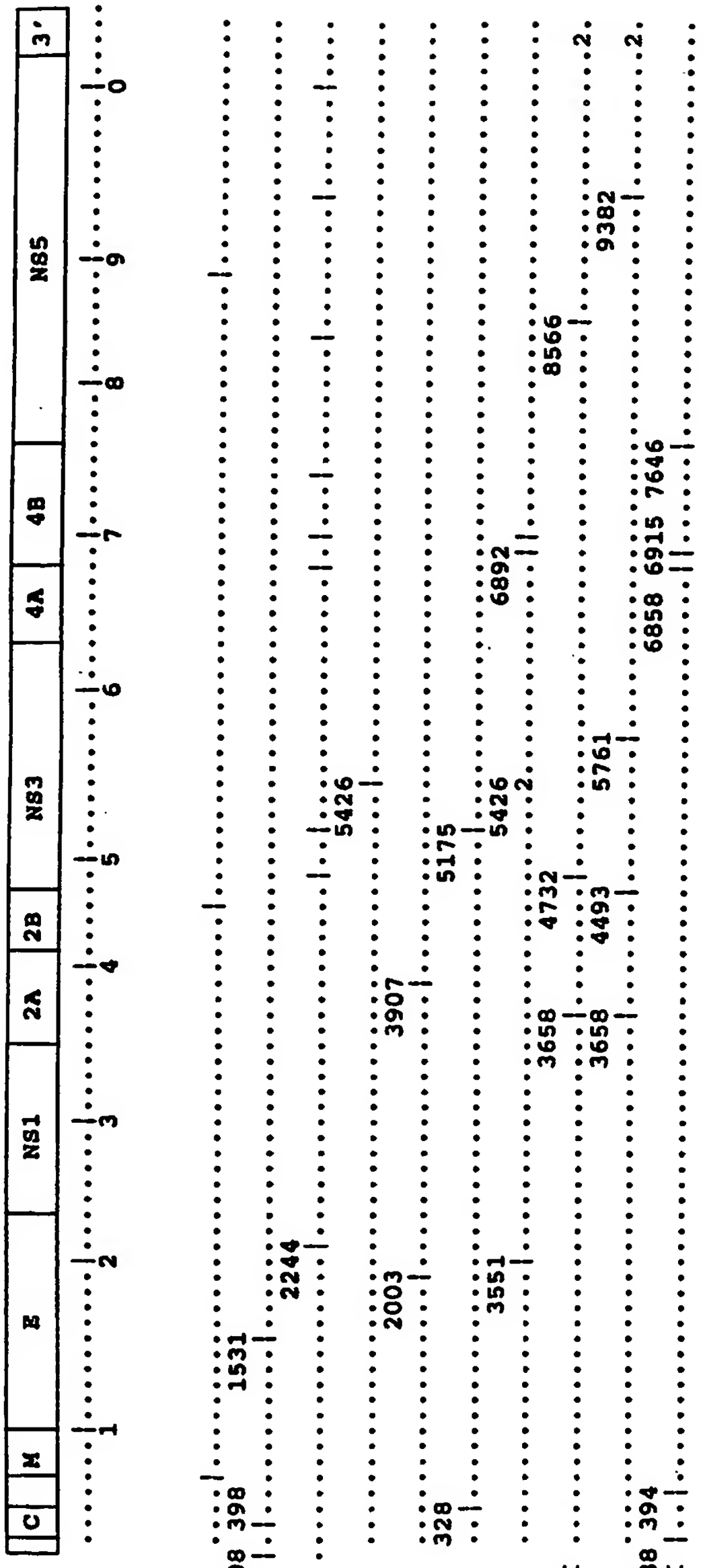

C
M
E
NS1
2A
2B
NS3
4A
4B
3'
1
2
3
4
5
6
7
8
9
0
PmlI 208
PstI 398
XcmI
XhoI
MunI 328
AseI
AvaI
BsaHI
BanI 88
BglII
1531
2244
2003
3551
3907
5175
5426
3658
4732
3658
4493
5761
6892
6858
6915
7646
8566
9382

FIGURE 3B

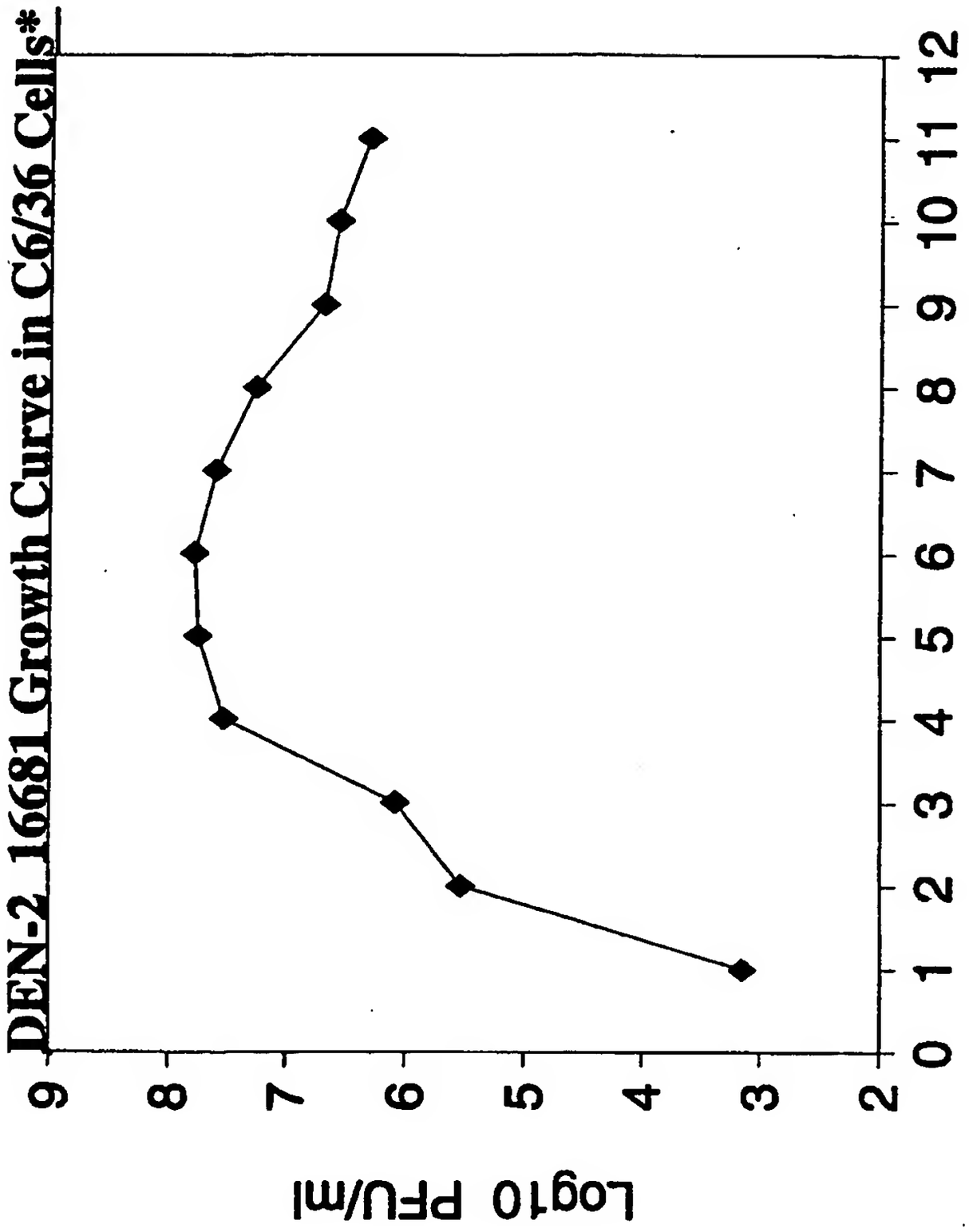

DEN-2_16681 Growth Curve in C6/36 Cells*
Log10 PFU/ml
9
8
7
6
5
4
3
2
0 1 2 3 4 5 6 7 8 9 10 11 12
Days Postinfection
* M.O.I. = 0.004, 150cm2, 40 ml

FIGURE 4

FIGURE 5A

FIGURE 5B

FIGURE 5C

FIGURE 5D

B

RT/PCR Amplicon	Expected Amplicon Length	Up-Amplimer	Down-Amplimer
T7-F1	1552-bp	D2-SMT71	cD2-1503
F2	3327-bp	D2-1261	cD2-4557
F2-Sal	2742-bp	D2-1261	cD2-4002
Sal-F2	2388-bp	D2-2170	cD2-4557
F3	2368-bp	D2-4257	cD2-6624
F4	3304-bp	D2-6493	cD2-9796
F5	1032-bp	D2-9656	cD2-10687.Xba

FIGURE 5E

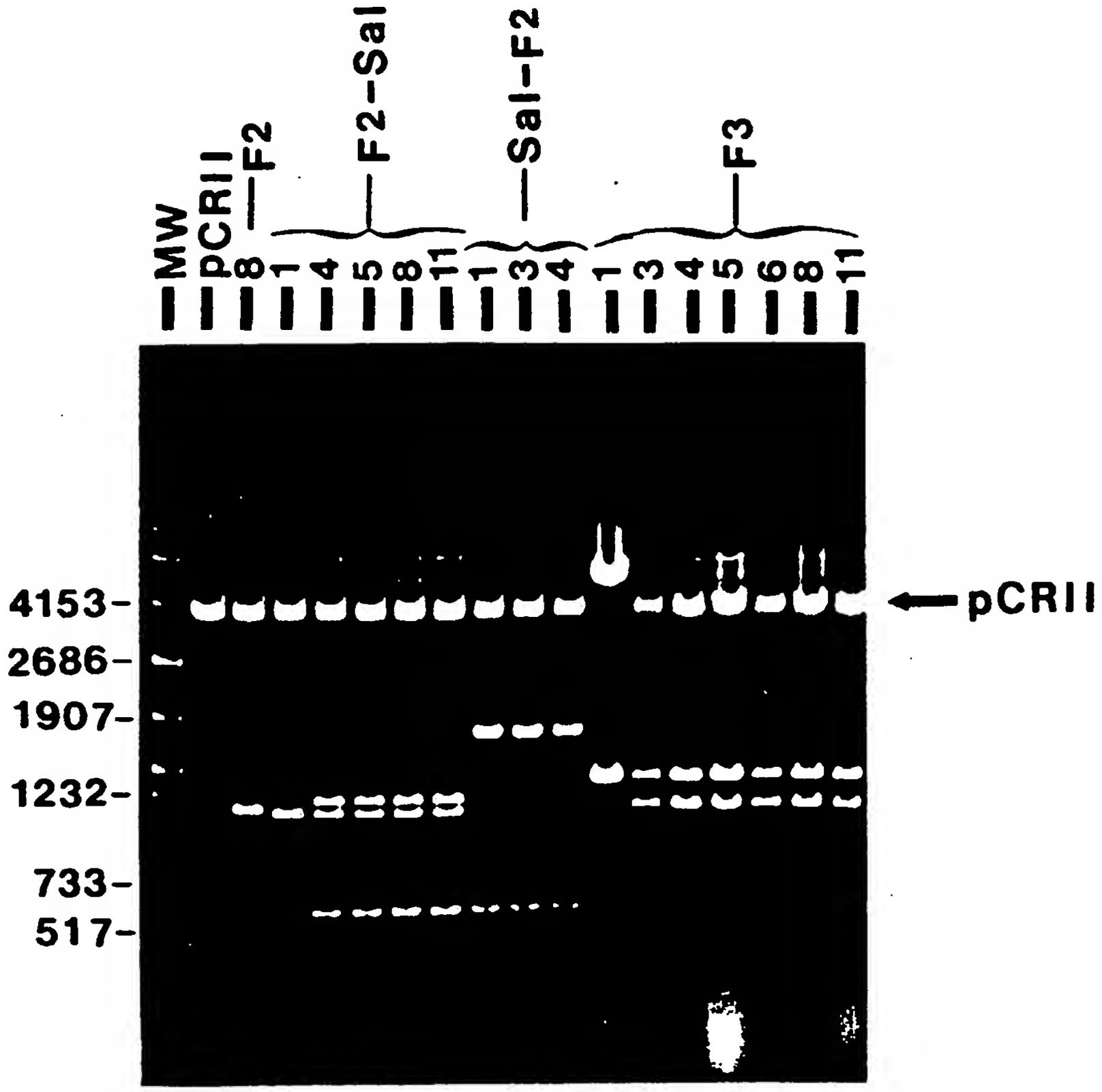

Plasmid Eco RI Digests
MW
PCRII
F2
8
F2
1
4
5
8
11
Sal I-F2
Sa
1
3
4
F3
1
3
4
5
6
8
11
4153-
2686-
1907-
1232-
733-
517-
pCRII

FIGURE 6

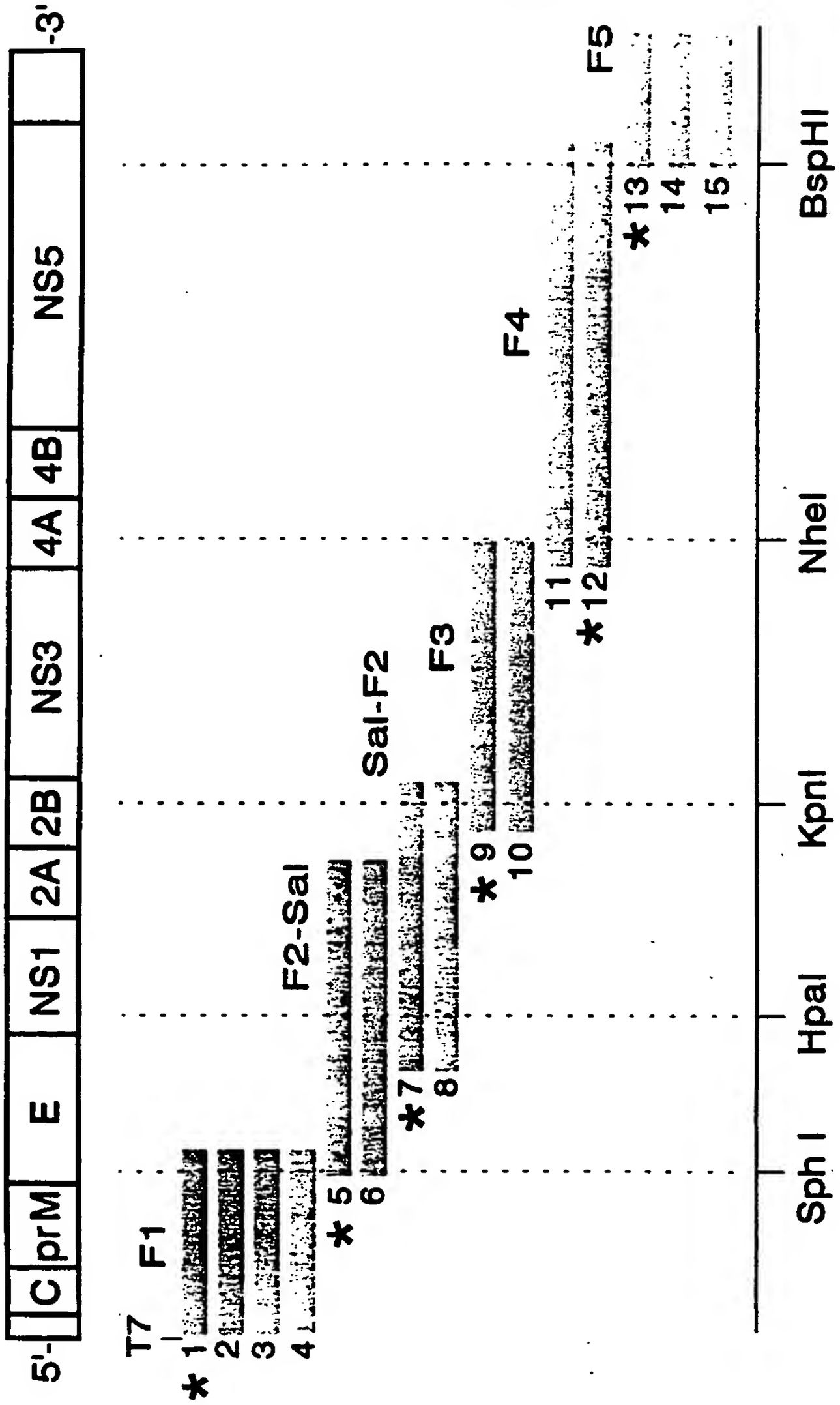

5'-
C
prM
E
NS1
2A
2B
NS3
4A
4B
NS5
-3'
T7
F1
F2-Sal
Sal-F2
F3
F4
F5
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
*
HpaI
KpnI
NheI
BspHI
Sph I

FIGURE 7

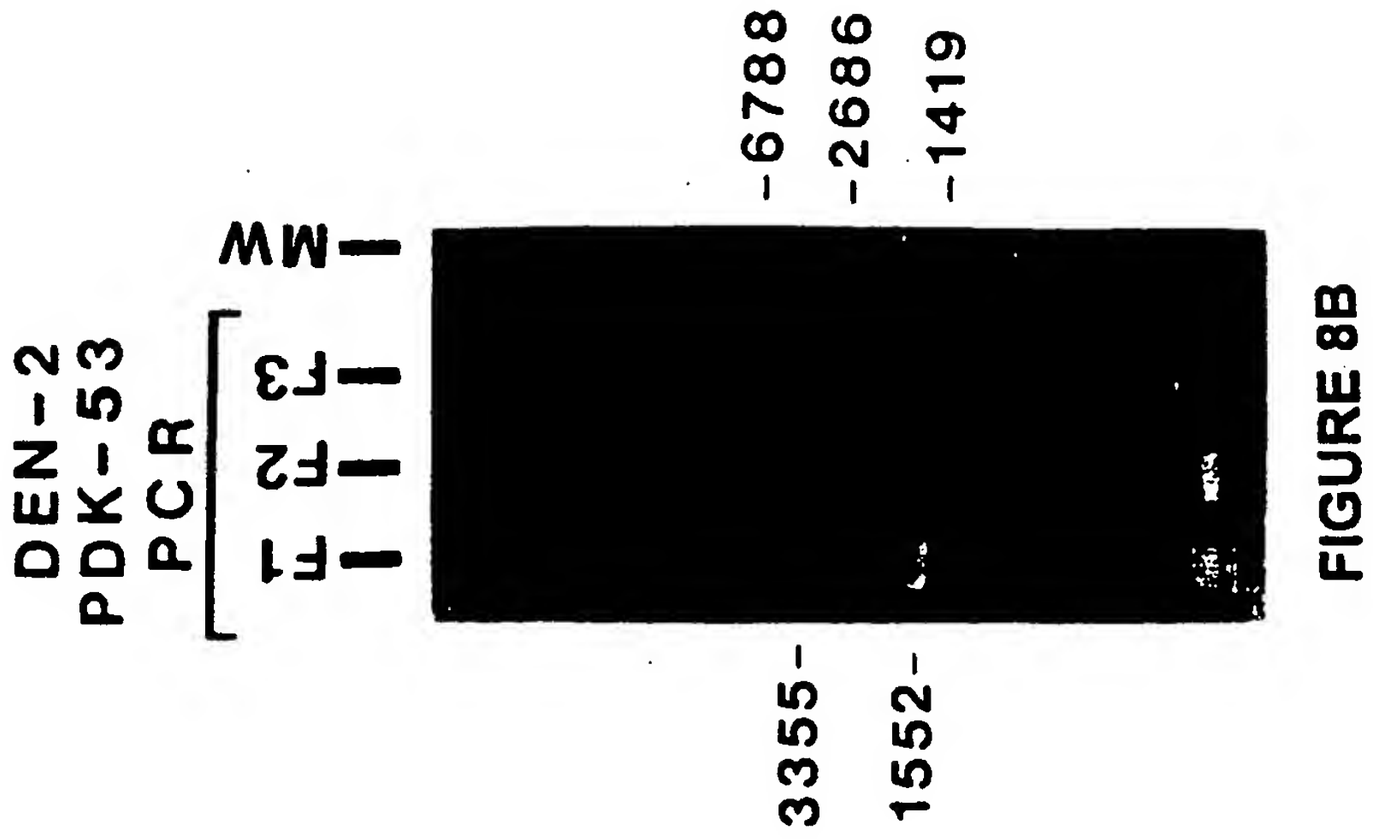


FIGURE 8B

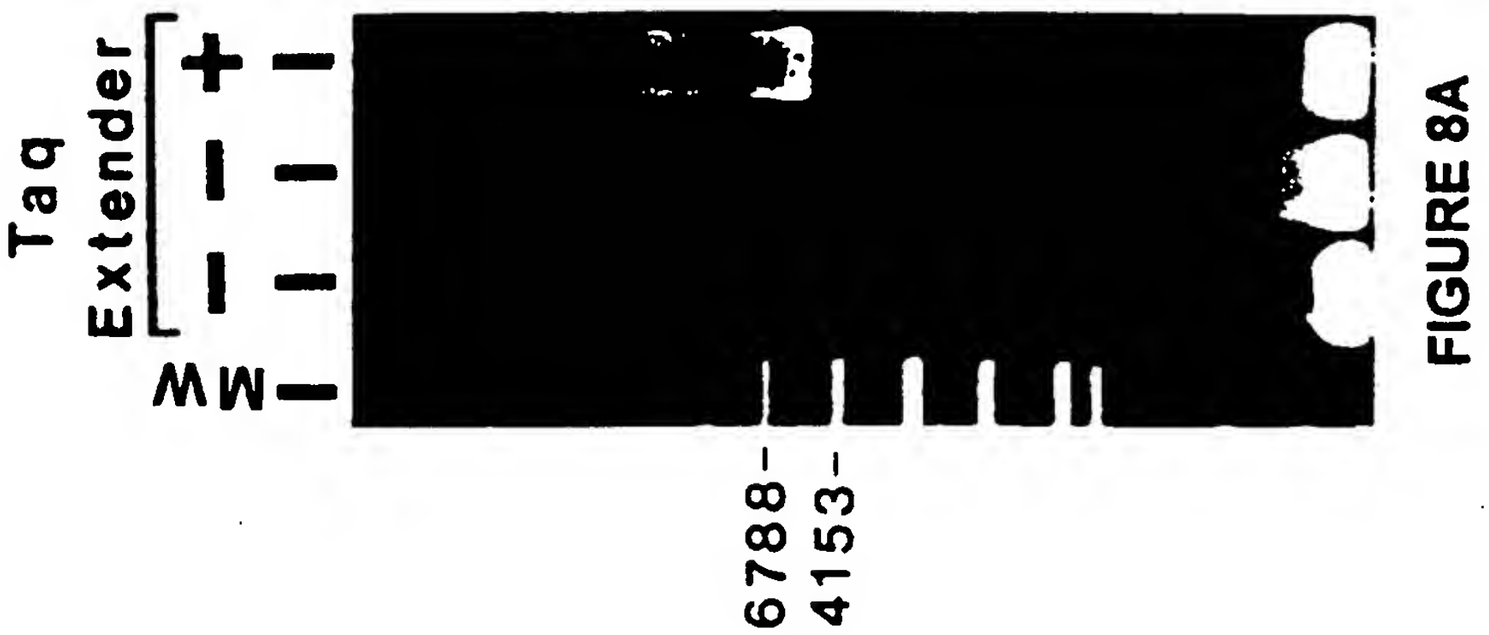


FIGURE 8A

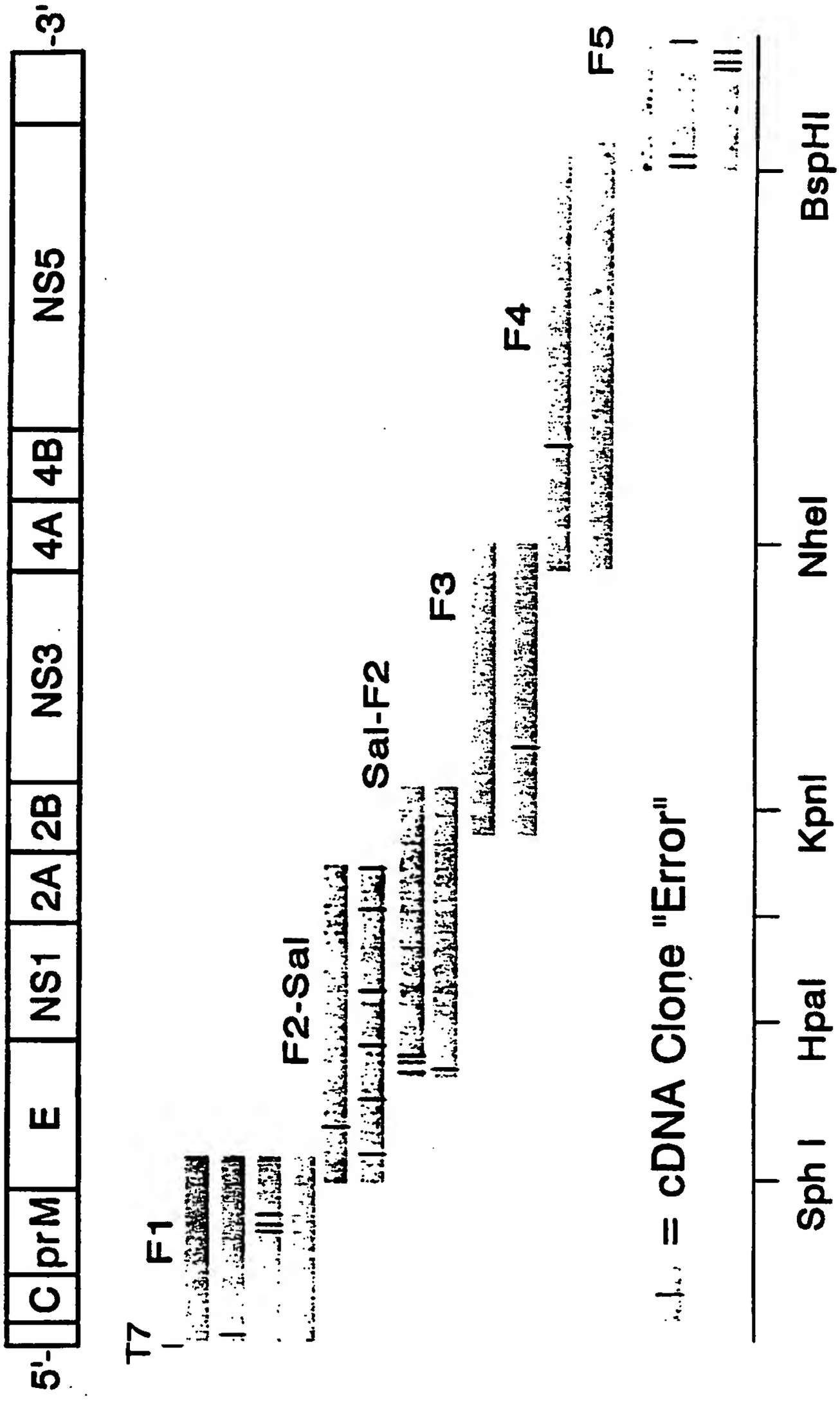

5'-
C
prM
E
NS1
2A
2B
NS3
4A
4B
NS5
-3'
T7
F1
F2-Sal
Sal-F2
F3
F4
F5
I
II
III
BspHI
Hpal
KpnI
NheI
Sph I
= cDNA Clone "Error"

FIGURE 9

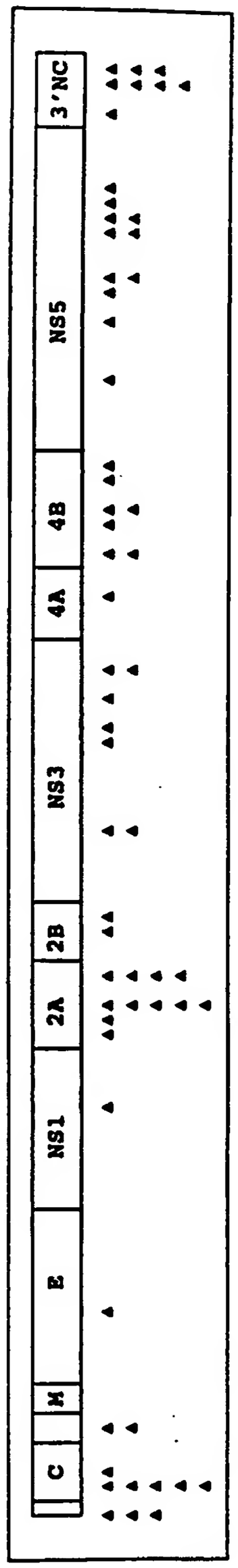


FIGURE 10

	5'-NC			Capsid						prM		E	NS1	NS2A							
												1	3	3	3	3	3	3	3	3	3
				1	1	2	2	2	3	6	7	4	0	6	6	7	8	8	8	8	9
	4	6	7	0	3	1	2	3	2	5	0	4	8	1	4	7	0	0	2	2	8
	9	9	7	3	2	0	6	6	1	1	3	5	1	1	5	6	8	9	2	4	7
DEN-2-16681.RK	A	A	A	A	T	G	T	T	A	G	G	T	T	C	G	A	C	T	C	T	A
DEN-2-16681.BLOK	G	T	G	G	A	A	C	A	C	C	C	C	C	T	A	C	T	C	T	C	G

	NS2A			NS2B		NS3							NS4A	NS4B							
	4	4	4	4	4	5	5	5	5	5	6	6	6	6	6	7	7	7	7	7
	0	0	0	2	3	0	0	6	7	9	1	1	5	8	8	1	1	2	4	5
	1	4	6	4	4	3	7	4	0	5	1	1	8	6	7	3	9	1	0	6
	8	4	2	1	4	4	6	9	5	5	3	4	0	4	0	9	6	3	3	1
DEN-2-16681.RK	C	A	A	T	T	C	G	G	T	T	A	A	T	C	G	T	G	A	C	A
DEN-2-16681.BLOK	T	G	G	C	C	T	A	C	C	C	C	C	C	A	A	C	C	T	T	G

	NS5										3'-NC							
											1	1	1	1	1	1	1	1
	8	8	9	9	9	9	9	9	9	9	0	0	0	0	0	0	0	0
	5	9	2	2	3	5	5	6	6	7	8	3	6	6	6	6	6	6
	4	3	0	2	0	9	9	2	2	3	4	3	3	4	5	5	5	7
	2	1	7	1	3	1	2	0	1	2	5	0	7	6	1	2	5	9	3
DEN-2-16681.RK	G	G	C	A	G	C	G	T	G	T	C	G	A	G	T	G	G	C	C
DEN-2-16681.BLOK	T	C	T	C	C	G	C	C	A	C	A	C	C	A	A	A	A	A	T

FIGURE 11

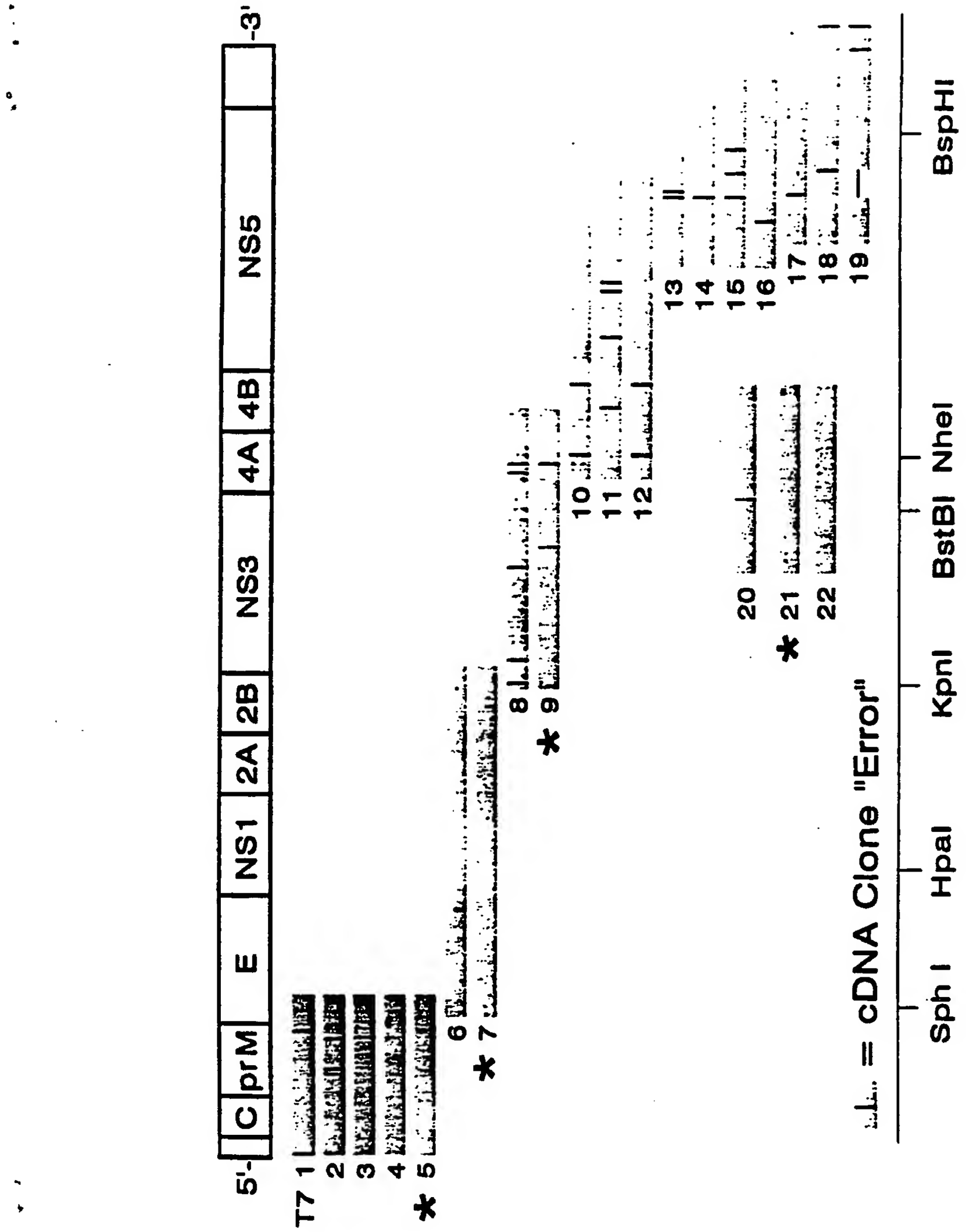

5'-
C
prM
E
NS1
2A
2B
NS3
4A
4B
-3'
T7 1
2
3
4
* 5
6
* 7
8
* 9
10
11
12
13
14
15
16
17
18
19
20
* 21
22
BspHI
HpaI
KpnI
BstBI
NheI
Sph I
= cDNA Clone "Error"

FIGURE 12

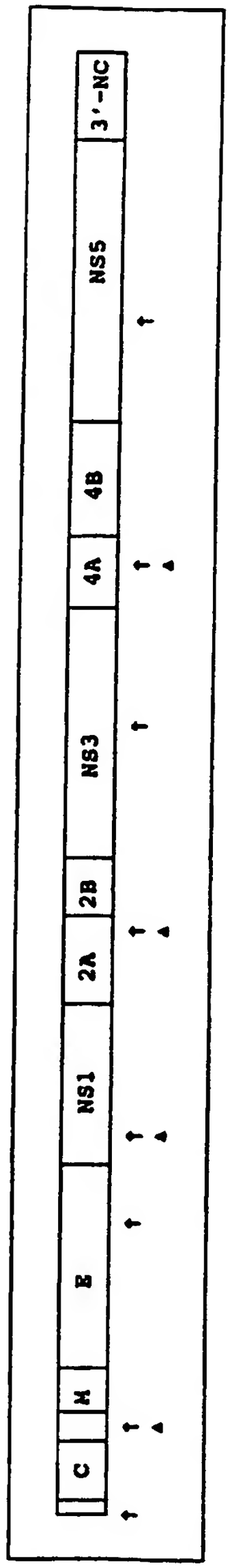


FIGURE 13

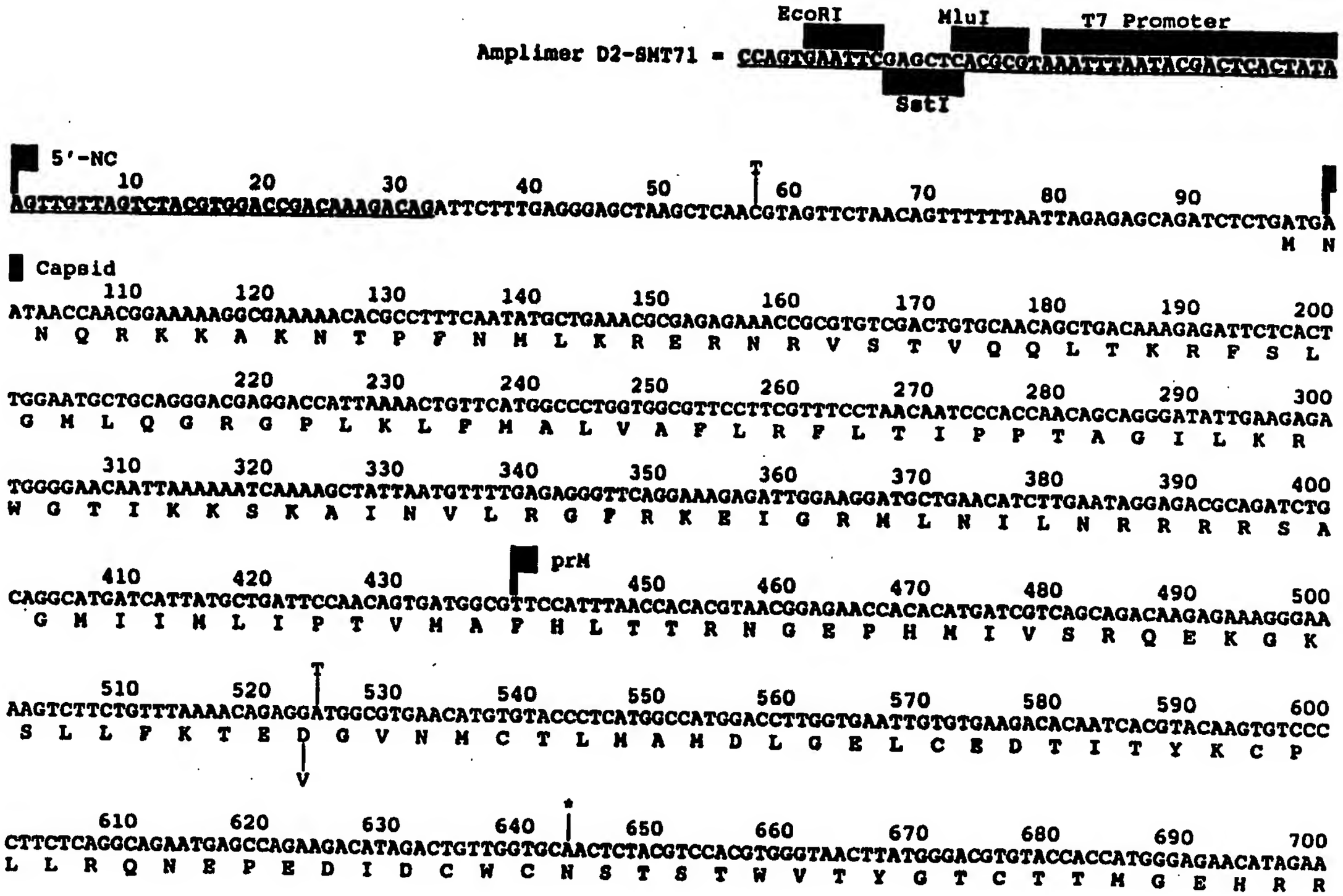


FIGURE 14A

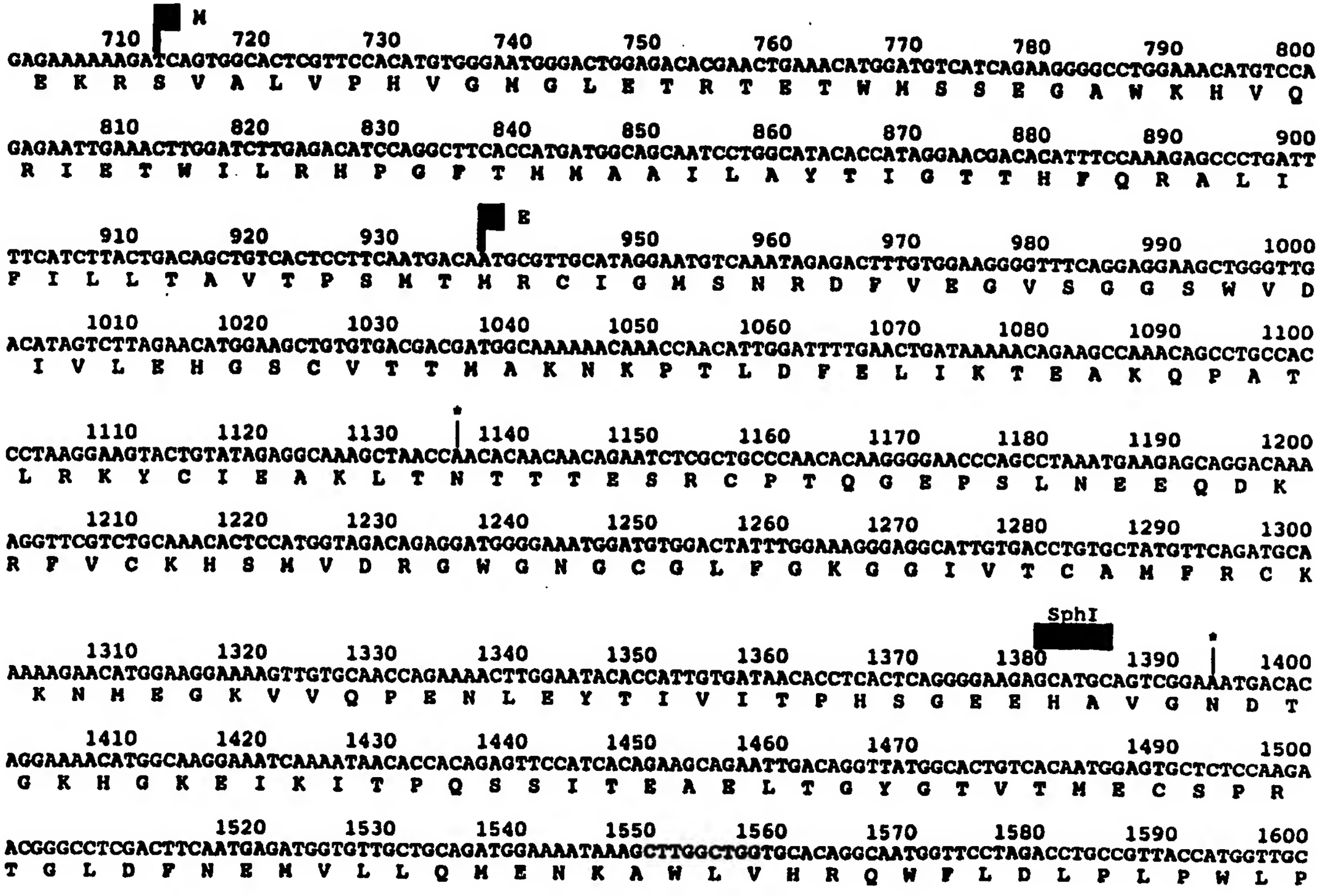

M
710 720 730 740 750 760 770 780 790 800
GAGAAAAAAGATCAGTGGCACTCGTTCCACATGTGGGAATGGGACTGGAGACACGAACTGAAACATGGATGTCATCAGAAGGGGCCTGGAAACATGTCCA
E K R S V A L V P H V G M G L E T R T E T W M S S E G A W K H V Q
810 820 830 840 850 860 870 880 890 900
GAGAATTGAAACTTGGATCTTGAGACATCCAGGCTTCACCATGATGGCAGCAATCCTGGCATACACCATAGGAACGACACATTTCCAAAGAGCCCTGATT
R I E T W I L R H P G F T M M A A I L A Y T I G T T H F Q R A L I
E
910 920 930 950 960 970 980 990 1000
TTCATCTTACTGACAGCTGTCACTCCTTCAATGACAATGCGTTGCATAGGAATGTCAAATAGAGACTTTGTGGAAGGGGTTTCAGGAGGAAGCTGGGTTG
F I L L T A V T P S M T M R C I G M S N R D F V E G V S G G S W V D
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
ACATAGTCTTAGAACATGGAAGCTGTGTGACGACGATGGCAAAAAACAAACCAACATTGGATTTTGAACTGATAAAAACAGAAGCCAAACAGCCTGCCAC
I V L E H G S C V T T M A K N K P T L D F E L I K T E A K Q P A T
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
CCTAAGGAAGTACTGTATAGAGGCAAAGCTAACCAACACAACAACAGAATCTCGCTGCCCAACACAAGGGGAACCCAGCCTAAATGAAGAGCAGGACAAA
L R K Y C I E A K L T N T T T E S R C P T Q G E P S L N E E Q D K
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
AGGTTCGTCTGCAAACACTCCATGGTAGACAGAGGATGGGGAAATGGATGTGGACTATTTGGAAAGGGAGGCATTGTGACCTGTGCTATGTTCAGATGCA
R F V C K H S M V D R G W G N G C G L F G K G G I V T C A M F R C K
SphI
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
AAAAGAACATGGAAGGAAAAGTTGTGCAACCAGAAAACTTGGAATACACCATTGTGATAACACCTCACTCAGGGGAAGAGCATGCAGTCGGAAATGACAC
K N M E G K V V Q P E N L E Y T I V I T P H S G E E H A V G N D T
1410 1420 1430 1440 1450 1460 1470 1490 1500
AGGAAAACATGGCAAGGAAATCAAAATAACACCACAGAGTTCCATCACAGAAGCAGAATTGACAGGTTATGGCACTGTCACAATGGAGTGCTCTCCAAGA
G K H G K E I K I T P Q S S I T E A E L T G Y G T V T M E C S P R
1520 1530 1540 1550 1560 1570 1580 1590 1600
ACGGGCCTCGACTTCAATGAGATGGTGTTGCTGCAGATGGAAAATAAAGCTTGGCTGGTGCACAGGCAATGGTTCCTAGACCTGCCGTTACCATGGTTGC
T G L D F N E M V L L Q M E N K A W L V H R Q W F L D L P L P W L P

FIGURE 14B

```
      1610       1620       1630       1640       1650       1660       1670       1680       1690       1700
CCGGAGCGGACACACAAGGGTCAAATTGGATACAGAAAGAGACATTGGTCACTTTCAAAAATCCCCATGCGAAGAAACAGGATGTTGTTGTTTTAGGATC
  G  A  D  T  Q  G  S  N  W  I  Q  K  E  T  L  V  T  F  K  N  P  H  A  K  K  Q  D  V  V  V  L  G  S

      1710       1720       1730       1740       1750       1760       1770       1780       1790       1800
CCAAGAAGGGGCCATGCACACAGCACTTACAGGGGCCACAGAAATCCAAATGTCATCAGGAAACTTACTCTTCACAGGACATCTCAAGTGCAGGCTGAGA
  Q  E  G  A  M  H  T  A  L  T  G  A  T  E  I  Q  M  S  S  G  N  L  L  F  T  G  H  L  K  C  R  L  R

      1810       1820       1830       1840       1850       1860       1870       1880       1890       1900
ATGGACAAGCTACAGCTCAAAGGAATGTCATACTCTATGTGCACAGGAAAGTTTAAAGTTGTGAAGGAAATAGCAGAAACACAACATGGAACAATAGTTA
M  D  K  L  Q  L  K  G  M  S  Y  S  M  C  T  G  K  F  K  V  V  K  E  I  A  E  T  Q  H  G  T  I  V  I

      1910       1920       1930       1940       1950       1960       1970       1980       1990       2000
TCAGAGTGCAATATGAAGGGGACGGCTCTCCATGCAAGATCCCTTTTGAGATAATGGATTTGGAAAAAAGACATGTCTTAGGTCGCCTGATTACAGTCAA
  R  V  Q  Y  E  G  D  G  S  P  C  K  I  P  F  E  I  M  D  L  E  K  R  H  V  L  G  R  L  I  T  V  N

                                                  T
      2010       2020       2030       2040       2050  |    2060       2070       2080       2090       2100
CCCAATTGTGACAGAAAAAGATAGCCCAGTCAACATAGAAGCAGAACCTCCATTCGGAGACAGCTACATCATCATAGGAGTAGAGCCGGGACAACTGAAG
  P  I  V  T  E  K  D  S  P  V  N  I  E  A  E  P  P  F  G  D  S  Y  I  I  I  G  V  E  P  G  Q  L  K

      2110       2120       2130       2140       2150       2160       2170       2180       2190       2200
CTCAACTGGTTTAAGAAAGGAAGTTCTATCGGCCAAATGTTTGAGACAACAATGAGGGGGGCGAAGAGAATGGCCATTTTAGGTGACACAGCCTGGGATT
L  N  W  F  K  K  G  S  S  I  G  Q  M  F  E  T  T  M  R  G  A  K  R  M  A  I  L  G  D  T  A  W  D  F

      2210       2220       2230       2240       2250       2260       2270       2280       2290       2300
TTGGATCCTTGGGAGGAGTGTTTACATCTATAGGAAAGGCTCTCCACCAAGTCTTTGGAGCAATCTATGGAGCTGCCTTCAGTGGGGTTTCATGGACTAT
  G  S  L  G  G  V  F  T  S  I  G  K  A  L  H  Q  V  F  G  A  I  Y  G  A  A  F  S  G  V  S  W  T  M

      2310       2320       2330       2340       2350       2360       2370       2380       2390       2400
GAAAATCCTCATAGGAGTCATTATCACATGGATAGGAATGAATTCACGCAGCACCTCACTGTCTGTGACACTAGTATTGGTGGGAATTGTGACACTGTAT
 K  I  L  I  G  V  I  I  T  W  I  G  M  N  S  R  S  T  S  L  S  V  T  L  V  L  V  G  I  V  T  L  Y

                                  NS1
      2410       2420  |    2430       2440       2450       2460       2470       2480       2490       2500
TTGGGAGTCATGGTGCAGGCCGATAGTGGTTGCGTTGTGAGCTGGAAAAACAAAGAACTGAAATGTGGCAGTGGGATTTTCATCACAGACAACGTGCACA
L  G  V  M  V  Q  A  D  S  G  C  V  V  S  W  K  N  K  E  L  K  C  G  S  G  I  F  I  T  D  N  V  H  T
```


FIGURE 14C

```
      2510       2520       2530       2540       2550       2560       2570                  A          2590       2600
CATGGACAGAACAATACAAGTTCCAACCAGAATCCCCTTCAAAACTAGCTTCAGCTATCCAGAAAGCCCATGAAGAGGGCATTTGTGGAATCCGCTCAGT
  W  T  E  Q  Y  K  F  Q  P  E  S  P  S  K  L  A  S  A  I  Q  K  A  H  E  E  G  I  C  G  I  R  S  V
                                                                                 D

                                                                                HpaI
      2610       2620       2630       2640       2650       2660       2670  ▮▮▮▮▮▮▮▮  2690       2700
AACAAGACTGGAGAATCTGATGTGGAAACAAATAACACCAGAATTGAATCACATTCTATCAGAAAATGAGGTGAAGTTAACTATTATGACAGGAGACATC
  T  R  L  E  N  L  M  W  K  Q  I  T  P  E  L  N  H  I  L  S  E  N  E  V  K  L  T  I  M  T  G  D  I

      2710       2720       2730       2740       2750       2760       2770       2780       2790       2800
AAAGGAATCATGCAGGCAGGAAAACGATCTCTGCGGCCTCAGCCCACTGAGCTGAAGTATTCATGGAAAACATGGGGCAAAGCAAAAATGCTCTCTACAG
  K  G  I  M  Q  A  G  K  R  S  L  R  P  Q  P  T  E  L  K  Y  S  W  K  T  W  G  K  A  K  M  L  S  T  E

           *
      2810       2820       2830       2840       2850       2860       2870       2880       2890       2900
AGTCTCATAACCAGACCTTTCTCATTGATGGCCCCGAAACAGCAGAATGCCCCAACACAAATAGAGCTTGGAATTCGTTGGAAGTTGAAGACTATGGCTT
  S  H  N  Q  T  F  L  I  D  G  P  E  T  A  E  C  P  N  T  N  R  A  W  N  S  L  E  V  E  D  Y  G  F

      2910       2920       2930       2940       2950       2960       2970       2980       2990       3000
TGGAGTATTCACCACCAATATATGGCTAAAATTGAAAGAAAAACAGGATGTATTCTGCGACTCAAAACTCATGTCAGCGGCCATAAAAGACAACAGAGCC
  G  V  F  T  T  N  I  W  L  K  K  L  K  E  K  Q  D  V  F  C  D  S  K  L  M  S  A  A  I  K  D  N  R  A

                                                  *
      3010       3020       3030                  3050       3060       3070       3080       3090       3100
GTCCATGCCGATATGGGTTATTGGATAGAAAGTGCACTCAATGACACATGGAAGATAGAGAAAGCCTCTTTCATTGAAGTTAAAAACTGCCACTGGCCAA
  V  H  A  D  M  G  Y  W  I  E  S  A  L  N  D  T  W  K  I  E  K  A  S  F  I  E  V  K  N  C  H  W  P  K

      3110       3120       3130       3140       3150       3160       3170       3180       3190       3200
AATCACACACCCTCTGGAGCAATGGAGTGCTAGAAAGTGAGATGATAATTCCAAAGAATCTCGCTGGACCAGTGTCTCAACACAACTATAGACCAGGCTA
  S  H  T  L  W  S  N  G  V  L  E  S  E  M  I  I  P  K  N  L  A  G  P  V  S  Q  H  N  Y  R  P  G  Y

      3210       3220       3230       3240       3250       3260       3270       3280       3290       3300
CCATACACAAATAACAGGACCATGGCATCTAGGTAAGCTTGAGATGGACTTTGATTTCTGTGATGGAACAACAGTGGTAGTGACTGAGGACTGCGGAAAT
  H  T  Q  I  T  G  P  W  H  L  G  K  L  E  M  D  F  D  F  C  D  G  T  T  V  V  V  T  E  D  C  G  N
```

FIGURE 14D

```
      3310       3320       3330       3340       3350       3360       3370       3380       3390       3400
AGAGGACCCTCTTTGAGAACAACCACTGCCTCTGGAAAACTCATAACAGAATGGTGCTGCCGATCTTGCACATTACCACCGCTAAGATACAGAGGTGAGG
R  G  P  S  L  R  T  T  T  A  S  G  K  L  I  T  E  W  C  C  R  S  C  T  L  P  P  L  R  Y  R  G  E  D

                                                                                  NS2A
      3410       3420       3430       3440       3450       3460       3470       3480       3490       3500
ATGGGTGCTGGTACGGGATGGAAATCAGACCATTGAAGGAGAAAGAAGAGAATTTGGTCAACTCCTTGGTCACAGCTGGACATGGGCAGGTCGACAACTT
  G  C  W  Y  G  M  E  I  R  P  L  K  E  K  E  E  N  L  V  N  S  L  V  T  A  G  H  G  Q  V  D  N  F

      3510       3520       3530       3540       3550       3560       3570       3580       3590       3600
TTCACTAGGAGTCTTGGGAATGGCATTGTTCCTGGAGGAAATGCTTAGGACCCGAGTAGGAACGAAACATGCAATACTACTAGTTGCAGTTTCTTTTGTG
 S  L  G  V  L  G  M  A  L  F  L  E  E  M  L  R  T  R  V  G  T  K  H  A  I  L  L  V  A  V  S  F  V

      3610       3620       3630       3640       3650       3660       3670       3680       3690       3700
ACATTGATCACAGGGAACATGTCCTTTAGAGACCTGGGAAGAGTGATGGTTATGGTAGGCGCCACTATGACGGATGACATAGGTATGGGCGTGACTTATC
T  L  I  T  G  N  M  S  F  R  D  L  G  R  V  M  V  M  V  G  A  T  M  T  D  D  I  G  M  G  V  T  Y  L

      3710       3720       3730       3740       3750       3760       3770       3780       3790       3800
TTGCCCTACTAGCAGCCTTCAAAGTCAGACCAACTTTTGCAGCTGGACTACTCTTGAGAAAGCTGACCTCCAAGGAATTGATGATGACTACTATAGGAAT
  A  L  L  A  A  F  K  V  R  P  T  F  A  A  G  L  L  L  R  K  L  T  S  K  E  L  M  M  T  T  I  G  I

      3810       3820       3830       3840       3850       3860       3870       3880       3890       3900
TGTACTCCTCTCCCAGAGCACCATACCAGAGACCATTCTTGAGTTGACTGATGCGTTAGCCTTAGGCATGATGGTCCTCAAAATGGTGAGAAATATGGAA
 V  L  L  S  Q  S  T  I  P  E  T  I  L  E  L  T  D  A  L  A  L  G  M  M  V  L  K  M  V  R  N  M  E

      3910       3920       3930       3940       3950       3960       3970       3980       3990       4000
AAGTATCAATTGGCAGTGACTATCATGGCTATCTTGTGCGTCCCAAACGCAGTGATATTACAAAACGCATGGAAAGTGAGTTGCACAATATTGGCAGTGG
K  Y  Q  L  A  V  T  I  M  A  I  L  C  V  P  N  A  V  I  L  Q  N  A  W  K  V  S  C  T  I  L  A  V  V

                 T
      4010       |  4030       4040       4050       4060       4070       4080       4090       4100
TGTCCGTTTCCCCACTGCTCTTAACATCCTCACAGCAAAAAACAGATTGGATACCATTAGCATTGACGATCAAAGGTCTCAATCCAACAGCTATTTTTCT
  S  V  S  P  L  L  L  T  S  S  Q  Q  K  T  D  W  I  P  L  A  L  T  I  K  G  L  N  P  T  A  I  F  L
                 |
                 P

                          NS2B
      4110       4120       4130       4140       4150       4160       4170       4180       4190       4200
AACAACCCTCTCAAGAACCAGCAAGAAAAGGAGCTGGCCATTAAATGAGGCTATCATGGCAGTCGGGATGGTGAGCATTTTAGCCAGTTCTCTCCTAAAA
 T  T  L  S  R  T  S  K  K  R  S  W  P  L  N  E  A  I  M  A  V  G  M  V  S  I  L  A  S  S  L  L  K
```

FIGURE 14E

```
      4210       4220       4230       4240       4250       4260       4270       4280       4290       4300
AATGATATTCCCATGACAGGACCATTAGTGGCTGGAGGGCTCCTCACTGTGTGCTACGTGCTCACTGGACGATCGGCCGATTTGGAACTGGAGAGAGCAG
N  D  I  P  M  T  G  P  L  V  A  G  G  L  L  T  V  C  Y  V  L  T  G  R  S  A  D  L  E  L  E  R  A  A

      4310       4320       4330       4340       4350       4360       4370       4380       4390       4400
CCGATGTCAAATGGGAAGACCAGGCAGAGATATCAGGAAGCAGTCCAATCCTGTCAATAACAATATCAGAAGATGGTAGCATGTCGATAAAAAATGAAGA
  D  V  K  W  E  D  Q  A  E  I  S  G  S  S  P  I  L  S  I  T  I  S  E  D  G  S  M  S  I  K  N  E  E

                                                                                                KpnI
      4410       4420       4430       4440       4450       4460       4470       4480       4490       4500
GGAAGAACAAACACTGACCATACTCATTAGAACAGGATTGCTGGTGATCTCAGGACTTTTTCCTGTATCAATACCAATCACGGCAGCAGCATGGTACCTG
E  E  Q  T  L  T  I  L  I  R  T  G  L  L  V  I  S  G  L  F  P  V  S  I  P  I  T  A  A  A  W  Y  L

                  NS3
      4510       4520       4530       4540       4550       4560       4570       4580       4590       4600
TGGGAAGTGAAGAAACAACGGGCCGGAGTATTGTGGGATGTTCCTTCACCCCCACCCATGGGAAAGGCTGAACTGGAAGATGGAGCCTATAGAATTAAGC
W  E  V  K  K  Q  R  A  G  V  L  W  D  V  P  S  P  P  P  M  G  K  A  E  L  E  D  G  A  Y  R  I  K  Q

      4610       4620       4630       4640       4650       4660       4670       4680       4690       4700
AAAAAGGGATTCTTGGATATTCCCAGATCGGAGCCGGAGTTTACAAAGAAGGAACATTCCATACAATGTGGCATGTCACACGTGGCGCTGTTCTAATGCA
  K  G  I  L  G  Y  S  Q  I  G  A  G  V  Y  K  E  G  T  F  H  T  M  W  H  V  T  R  G  A  V  L  M  H

      4710       4720       4730       4740       4750       4760       4770       4780       4790       4800
TAAAGGAAAGAGGATTGAACCATCATGGGCGGACGTCAAGAAAGACCTAATATCATATGGAGGAGGCTGGAAGTTAGAAGGAGAATGGAAGGAAGGAGAA
K  G  K  R  I  E  P  S  W  A  D  V  K  K  D  L  I  S  Y  G  G  G  W  K  L  E  G  E  W  K  E  G  E

      4810       4820       4830       4840       4850       4860       4870       4880       4890       4900
GAAGTCCAGGTATTGGCACTGGAGCCTGGAAAAAATCCAAGAGCCGTCCAAACGAAACCTGGTCTTTTCAAAACCAACGCCGGAACAATAGGTGCTGTAT
E  V  Q  V  L  A  L  E  P  G  K  N  P  R  A  V  Q  T  K  P  G  L  F  K  T  N  A  G  T  I  G  A  V  S

      4910       4920       4930       4940       4950       4960       4970       4980       4990       5000
CTCTGGACTTTTCTCCTGGAACGTCAGGATCTCCAATTATCGACAAAAAAGGAAAAGTTGTGGGTCTTTATGGTAATGGTGTTGTTACAAGGAGTGGAGC
  L  D  F  S  P  G  T  S  G  S  P  I  I  D  K  K  G  K  V  V  G  L  Y  G  N  G  V  V  T  R  S  G  A

      5010       5020       5030       5040       5050       5060       5070       5080       5090       5100
ATATGTGAGTGCTATAGCCCAGACTGAAAAAAGCATTGAAGACAACCCAGAGATCGAAGATGACATTTTCCGAAAGAGAAGACTGACCATCATGGACCTC
Y  V  S  A  I  A  Q  T  E  K  S  I  E  D  N  P  E  I  E  D  D  I  F  R  K  R  R  L  T  I  M  D  L
```

FIGURE 14F

```
      5110       5120       5130       5140       5150       5160       5170       5180       5190       5200
CACCCAGGAGCGGGAAAGACGAAGAGATACCTTCCGGCCATAGTCAGAGAAGCTATAAAACGGGGTTTGAGAACATTAATCTTGGCCCCCACTAGAGTTG
H  P  G  A  G  K  T  K  R  Y  L  P  A  I  V  R  E  A  I  K  R  G  L  R  T  L  I  L  A  P  T  R  V  V

      5210       5220       5230       5240       5250       5260       5270       5280       5290       5300
TGGCAGCTGAAATGGAGGAAGCCCTTAGAGGACTTCCAATAAGATACCAGACCCCAGCCATCAGAGCTGAGCACACCGGGCGGGAGATTGTGGACCTAAT
  A  A  E  M  E  E  A  L  R  G  L  P  I  R  Y  Q  T  P  A  I  R  A  E  H  T  G  R  E  I  V  D  L  M

      5310       5320       5330       5340       5350       5360       5370       5380       5390       5400
GTGTCATGCCACATTTACCATGAGGCTGCTATCACCAGTTAGAGTGCCAAACTACAACCTGATTATCATGGACGAAGCCCATTTCACAGACCCAGCAAGT
 C  H  A  T  F  T  M  R  L  L  S  P  V  R  V  P  N  Y  N  L  I  I  M  D  E  A  H  F  T  D  P  A  S

      5410       5420       5430       5440       5450       5460       5470       5480       5490       5500
ATAGCAGCTAGAGGATACATCTCAACTCGAGTGGAGATGGGTGAGGCAGCTGGGATTTTTATGACAGCCACTCCCCCGGGAAGCAGAGACCCATTTCCTC
I  A  A  R  G  Y  I  S  T  R  V  E  M  G  E  A  A  G  I  F  M  T  A  T  P  P  G  S  R  D  P  F  P  Q

                                                  T
      5510       5520       5530       5540       |  5560       5570       5580       5590       5600
AGAGCAATGCACCAATCATAGATGAAGAAAGAGAAATCCCTGAACGCTCGTGGAATTCCGGACATGAATGGGTCACGGATTTTAAAGGGAAGACTGTTTG
  S  N  A  P  I  I  D  E  E  R  E  I  P  E  R  S  W  N  S  G  H  E  W  V  T  D  F  K  G  K  T  V  W

      5610       5620       5630       5640       5650       5660       5670       5680       5690       5700
GTTCGTTCCAAGTATAAAAGCAGGAAATGATATAGCAGCTTGCCTGAGGAAAAATGGAAAGAAAGTGATACAACTCAGTAGGAAGACCTTTGATTCTGAG
 F  V  P  S  I  K  A  G  N  D  I  A  A  C  L  R  K  N  G  K  K  V  I  Q  L  S  R  K  T  F  D  S  E

      5710       5720       5730       5740       5750       5760       5770       5780       5790       5800
TATGTCAAGACTAGAACCAATGATTGGGACTTCGTGGTTACAACTGACATTTCAGAAATGGGTGCCAATTTCAAGGCTGAGAGGGTTATAGACCCCAGAC
Y  V  K  T  R  T  N  D  W  D  F  V  V  T  T  D  I  S  E  M  G  A  N  F  K  A  E  R  V  I  D  P  R  R

      5810       5820       5830       5840       5850       5860       5870       5880       5890       5900
GCTGCATGAAACCAGTCATACTAACAGATGGTGAAGAGCGGGTGATTCTGGCAGGACCTATGCCAGTGACCCACTCTAGTGCAGCACAAAGAAGAGGGAG
  C  M  K  P  V  I  L  T  D  G  E  E  R  V  I  L  A  G  P  M  P  V  T  H  S  S  A  A  Q  R  R  G  R

      5910       5920       5930       5940       5950       5960       5970       5980       5990       6000
AATAGGAAGAAATCCAAAAAATGAGAATGACCAGTACATATACATGGGGGAACCTCTGGAAAATGATGAAGACTGTGCACACTGGAAAGAAGCTAAAATG
 I  G  R  N  P  K  N  E  N  D  Q  Y  I  Y  M  G  E  P  L  E  N  D  E  D  C  A  H  W  K  E  A  K  M

      6010       6020       6030       6040       6050       6060       6070       6080       6090       6100
CTCCTAGATAACATCAACACGCCAGAAGGAATCATTCCTAGCATGTTCGAACCAGAGCGTGAAAAGGTGGATGCCATTGATGGCGAATACCGCTTGAGAG
L  L  D  N  I  N  T  P  E  G  I  I  P  S  M  F  E  P  E  R  E  K  V  D  A  I  D  G  E  Y  R  L  R  G
```

FIGURE 14G

```
      6110       6120       6130       6140       6150       6160       6170       6180       6190       6200
GAGAAGCAAGGAAAACCTTTGTAGACTTAATGAGAAGAGGAGACCTACCAGTCTGGTTGGCCTACAGAGTGGCAGCTGAAGGCATCAACTACGCAGACAG
  E  A  R  K  T  F  V  D  L  M  R  R  G  D  L  P  V  W  L  A  Y  R  V  A  A  E  G  I  N  Y  A  D  R

      6210       6220       6230       6240       6250       6260       6270       6280       6290       6300
AAGGTGGTGTTTTGATGGAGTCAAGAACAACCAAATCCTAGAAGAAAACGTGGAAGTTGAAATCTGGACAAAAGAAGGGGAAAGGAAGAAATTGAAACCC
 R  W  C  F  D  G  V  K  N  N  Q  I  L  E  E  N  V  E  V  E  I  W  T  K  E  G  E  R  K  K  L  K  P

                                                                                 NS4A
      6310       6320       6330       6340       6350       6360       6370       6390       6400
AGATGGTTGGATGCTAGGATCTATTCTGACCCACTGGCGCTAAAAGAATTTAAGGAATTTGCAGCCGGAAGAAAGTCTCTGACCCTGAACCTAATCACAG
R  W  L  D  A  R  I  Y  S  D  P  L  A  L  K  E  F  K  E  F  A  A  G  R  K  S  L  T  L  N  L  I  T  E

      6410       6420       6430       6440       6450       6460       6470       6480       6490       6500
AAATGGGTAGGCTCCCAACCTTCATGACTCAGAAGGCAAGAGACGCACTGGACAACTTAGCAGTGCTGCACACGGCTGAGGCAGGTGGAAGGGCGTACAA
  M  G  R  L  P  T  F  M  T  Q  K  A  R  D  A  L  D  N  L  A  V  L  H  T  A  E  A  G  G  R  A  Y  N

                                                                                                      C
      6510       6520       6530       6540       6550       6560       6570       6580       6590       |
CCATGCTCTCAGTGAACTGCCGGAGACCCTGGAGACATTGCTTTTACTGACACTTCTGGCTACAGTCACGGGAGGGATCTTTTTATTCTTGATGAGCGGA
 H  A  L  S  E  L  P  E  T  L  E  T  L  L  L  L  T  L  L  A  T  V  T  G  G  I  F  L  F  L  M  S  G
                                                                                                      |
                                                                                                      A

                                  NheI
      6610       6620       6630       6640       6650       6660       6670       6680       6690       6700
AGGGGCATAGGGAAGATGACCCTGGGAATGTGCTGCATAATCACGGCTAGCATCCTCCTATGGTACGCACAAATACAGCCACACTGGATAGCAGCTTCAA
R  G  I  G  K  M  T  L  G  M  C  C  I  I  T  A  S  I  L  L  W  Y  A  Q  I  Q  P  H  W  I  A  A  S  I

      6710       6720       6730       6740       6750       6760       6770       6780       6790       6800
TAATACTGGAGTTTTTTCTCATAGTTTTGCTTATTCCAGAACCTGAAAAACAGAGAACACCCCAAGACAACCAACTGACCTACGTTGTCATAGCCATCCT
  I  L  E  F  F  L  I  V  L  L  I  P  E  P  E  K  Q  R  T  P  Q  D  N  Q  L  T  Y  V  V  I  A  I  L

                 NS4B
      6810       6820       6830       6840       6850       6860       6870       6880       6890       6900
CACAGTGGTGGCCGCAACCATGGCAAACGAGATGGGTTTCCTAGAAAAAACGAAGAAAGATCTCGGATTGGGAAGCATTGCAACCCAGCAACCCGAGAGC
 T  V  V  A  A  T  M  A  N  E  M  G  F  L  E  K  T  K  K  D  L  G  L  G  S  I  A  T  Q  Q  P  E  S

      6910       6920       6930       6940       6950       6960       6970       6980       6990       7000
AACATCCTGGACATAGATCTACGTCCTGCATCAGCATGGACGCTGTATGCCGTGGCCACAACATTTGTTACACCAATGTTGAGACATAGCATTGAAAATT
N  I  L  D  I  D  L  R  P  A  S  A  W  T  L  Y  A  V  A  T  T  F  V  T  P  M  L  R  H  S  I  E  N  S
```

FIGURE 14H

```
      7010       7020       7030       7040       7050       7060       7070       7080       7090       7100
CCTCAGTGAATGTGTCCCTAACAGCTATAGCCAACCAAGCCACAGTGTTAATGGGTCTCGGGAAAGGATGGCCATTGTCAAAGATGGACATCGGAGTTCC
  S  V  N  V  S  L  T  A  I  A  N  Q  A  T  V  L  H  G  L  G  K  G  W  P  L  S  K  M  D  I  G  V  P

      7110       7120       7130       7140       7150       7160       7170       7180       7190       7200
CCTTCTCGCCATTGGATGCTACTCACAAGTCAACCCCATAACTCTCACAGCAGCTCTTTTCTTATTGGTAGCACATTATGCCATCATAGGGCCAGGACTC
  L  L  A  I  G  C  Y  S  Q  V  N  P  I  T  L  T  A  A  L  F  L  L  V  A  H  Y  A  I  I  G  P  G  L

      7210       7220       7230       7240       7250       7260       7270       7280       7290       7300
CAAGCAAAAGCAACCAGAGAAGCTCAGAAAAGAGCAGCGGCGGGCATCATGAAAAACCCAACTGTCGATGGAATAACAGTGATTGACCTAGATCCAATAC
  Q  A  K  A  T  R  E  A  Q  K  R  A  A  A  G  I  M  K  N  P  T  V  D  G  I  T  V  I  D  L  D  P  I  P

      7310       7320       7330       7340       7350       7360       7370       7380       7390       7400
CTTATGATCCAAAGTTTGAAAAGCAGTTGGGACAAGTAATGCTCCTAGTCCTCTGCGTGACTCAAGTATTGATGATGAGGACTACATGGGCTCTGTGTGA
  Y  D  P  K  F  E  K  Q  L  G  Q  V  M  L  L  V  L  C  V  T  Q  V  L  M  M  R  T  T  W  A  L  C  E

      7410       7420       7430                  7450       7460       7470       7480       7490       7500
GGCTTTAACCTTAGCTACCGGGCCCATCTCCACATTGTGGGAAGGAAATCCAGGGAGGTTTTGGAACACTACCATTGCGGTGTCAATGGCTAACATTTTT
  A  L  T  L  A  T  G  P  I  S  T  L  W  E  G  N  P  G  R  F  W  N  T  T  I  A  V  S  M  A  N  I  F

                                                                                  NS5
      7510       7520       7530       7540       7550                                     7590       7600
AGAGGGAGTTACTTGGCCGGAGCTGGACTTCTCTTTTCTATTATGAAGAACACAACCAACACAAGAAGGGGAACTGGCAACATAGGAGAGACGCTTGGAG
  R  G  S  Y  L  A  G  A  G  L  L  F  S  I  M  K  N  T  T  N  T  R  R  G  T  G  N  I  G  E  T  L  G  E

      7610       7620       7630       7640       7650       7660       7670       7680       7690       7700
AGAAATGGAAAAGCCGATTGAACGCATTGGGAAAAAGTGAATTCCAGATCTACAAGAAAAGTGGAATCCAGGAAGTGGATAGAACCTTAGCAAAAGAAGG
  K  W  K  S  R  L  N  A  L  G  K  S  E  F  Q  I  Y  K  K  S  G  I  Q  E  V  D  R  T  L  A  K  E  G

      7710       7720       7730       7740       7750       7760       7770       7780       7790       7800
CATTAAAAGAGGAGAAACGGACCATCACGCTGTGTCGCGAGGCTCAGCAAAACTGAGATGGTTCGTTGAGAGAAACATGGTCACACCAGAAGGGAAAGTA
  I  K  R  G  E  T  D  H  H  A  V  S  R  G  S  A  K  L  R  W  F  V  E  R  N  M  V  T  P  E  G  K  V

      7810       7820       7830       7840       7850       7860       7870       7880       7890       7900
GTGGACCTCGGTTGTGGCAGAGGAGGCTGGTCATACTATTGTGGAGGACTAAAGAATGTAAGAGAAGTCAAAGGCCTAACAAAAGGAGGACCAGGACACG
  V  D  L  G  C  G  R  G  G  W  S  Y  Y  C  G  G  L  K  N  V  R  E  V  K  G  L  T  K  G  G  P  G  H  E

      7910       7920       7930       7940       7950       7960       7970       7980       7990       8000
AAGAACCCATCCCCATGTCAACATATGGGTGGAATCTAGTGCGTCTTCAAAGTGGAGTTGACGTTTTCTTCATCCCGCCAGAAAAGTGTGACACATTATT
  E  P  I  P  M  S  T  Y  G  W  N  L  V  R  L  Q  S  G  V  D  V  F  F  I  P  P  E  K  C  D  T  L  L
```

FIGURE 14I

```
      8010       8020       8030       8040       8050       8060       8070       8080       8090       8100
GTGTGACATAGGGGAGTCATCACCAAATCCCACAGTGGAAGCAGGACGAACACTCAGAGTCCTTAACTTAGTAGAAAATTGGTTGAACAACAACACTCAA
 C  D  I  G  E  S  S  P  N  P  T  V  E  A  G  R  T  L  R  V  L  N  L  V  E  N  W  L  N  N  N  T  Q

      8110       8120       8130       8140       8150       8160       8170       8180       8190       8200
TTTTGCATAAAGGTTCTCAACCCATATATGCCCTCAGTCATAGAAAAAATGGAAGCACTACAAAGGAAATATGGAGGAGCCTTAGTGAGGAATCCACTCT
F  C  I  K  V  L  N  P  Y  M  P  S  V  I  E  K  M  E  A  L  Q  R  K  Y  G  G  A  L  V  R  N  P  L  S

      8210       8220       8230       8240       8250       8260       8270       8280       8290       8300
CACGAAACTCCACACATGAGATGTACTGGGTATCCAATGCTTCCGGGAACATAGTGTCATCAGTGAACATGATTTCAAGGATGTTGATCAACAGATTTAC
  R  N  S  T  H  E  M  Y  W  V  S  N  A  S  G  N  I  V  S  S  V  N  M  I  S  R  M  L  I  N  R  F  T

      8310       8320       8330       8340       8350       8360       8370       8380       8390       8400
AATGAGATACAAGAAAGCCACTTACGAGCCGGATGTTGACCTCGGAAGCGGAACCCGTAACATCGGGATTGAAAGTGAGATACCAAACCTAGATATAATT
 M  R  Y  K  K  A  T  Y  E  P  D  V  D  L  G  S  G  T  R  N  I  G  I  E  S  E  I  P  N  L  D  I  I

      8410       8420       8430       8440       8450       8460       8470       8480       8490       8500
GGGAAAAGAATAGAAAAAATAAAGCAAGAGCATGAAACATCATGGCACTATGACCAAGACCACCCATACAAAACGTGGGCATACCATGGTAGCTATGAAA
G  K  R  I  E  K  I  K  Q  E  H  E  T  S  W  H  Y  D  Q  D  H  P  Y  K  T  W  A  Y  H  G  S  Y  E  T

                                                                    T
      8510       8520       8530       8540       8550       8560                  8580       8590       8600
CAAAACAGACTGGATCAGCATCATCCATGGTCAACGGAGTGGTCAGGCTGCTGACAAAACCTTGGGACGTCGTCCCCATGGTGACACAGATGGCAATGAC
  K  Q  T  G  S  A  S  S  M  V  N  G  V  V  R  L  L  T  K  P  W  D  V  V  P  M  V  T  Q  M  A  M  T

      8610       8620       8630       8640       8650       8660       8670       8680       8690       8700
AGACACGACTCCATTTGGACAACAGCGCGTTTTTAAAGAGAAAGTGGACACGAGAACCCAAGAACCGAAAGAAGGCACGAAGAAACTAATGAAAATAACA
 D  T  T  P  F  G  Q  Q  R  V  F  K  E  K  V  D  T  R  T  Q  E  P  K  E  G  T  K  K  L  M  K  I  T

      8710       8720       8730       8740       8750       8760       8770       8780       8790       8800
GCAGAGTGGCTTTGGAAAGAATTAGGGAAGAAAAAGACACCCAGGATGTGCACCAGAGAAGAATTCACAAGAAAGGTGAGAAGCAATGCAGCCTTGGGGG
A  E  W  L  W  K  E  L  G  K  K  K  T  P  R  M  C  T  R  E  E  F  T  R  K  V  R  S  N  A  A  L  G  A

      8810       8820       8830       8840       8850       8860       8870       8880       8890       8900
CCATATTCACTGATGAGAACAAGTGGAAGTCGGCACGTGAGGCTGTTGAAGATAGTAGGTTTTGGGAGCTGGTTGACAAGGAAAGGAATCTCCATCTTGA
 I  F  T  D  E  N  K  W  K  S  A  R  E  A  V  E  D  S  R  F  W  E  L  V  D  K  E  R  N  L  H  L  E

      8910       8920       8930       8940       8950       8960       8970       8980       8990       9000
AGGAAAGTGTGAAACATGTGTGTACAACATGATGGGAAAAAGAGAGAAGAAGCTAGGGGAATTCGGCAAGGCAAAAGGCAGCAGAGCCATATGGTACATG
G  K  C  E  T  C  V  Y  N  M  M  G  K  R  E  K  K  L  G  E  F  G  K  A  K  G  S  R  A  I  W  Y  M
```

FIGURE 14J

```
      9010       9020       9030       9040       9050       9060       9070       9080       9090       9100
TGGCTTGGAGCACGCTTCTTAGAGTTTGAAGCCCTAGGATTCTTAAATGAAGATCACTGGTTCTCCAGAGAGAACTCCCTGAGTGGAGTGGAAGGAGAAG
W  L  G  A  R  F  L  E  F  E  A  L  G  F  L  N  E  D  H  W  F  S  R  E  N  S  L  S  G  V  E  G  E  G

      9110       9120       9130       9140       9150       9160       9170       9180       9190       9200
GGCTGCACAAGCTAGGTTACATTCTAAGAGACGTGAGCAAGAAAGAGGGAGGAGCAATGTATGCCGATGACACCGCAGGATGGGATACAAGAATCACACT
   L  H  K  L  G  Y  I  L  R  D  V  S  K  K  E  G  G  A  M  Y  A  D  D  T  A  G  W  D  T  R  I  T  L

      9210       9220       9230       9240       9250       9260       9270       9280       9290       9300
AGAAGACKKAAAAAATGAAGAAATGGTAACAAACCACATGGAAGGAGAACACAAGAAACTAGCCGAGGCCATTTTCAAACTAACGTACCAAAACAAGGTG
 E  D  ?  K  N  E  E  M  V  T  N  H  M  E  G  E  H  K  K  L  A  E  A  I  F  K  L  T  Y  Q  N  K  V

      9310       9320       9330       9340       9350       9360       9370       9380       9390       9400
GTGCGTGTGCAAAGACCAACACCAAGAGGCACAGTAATGGACATCATATCGAGAAGAGACCAAAGAGGTAGTGGACAAGTTGGCACCTATGGACTCAATA
V  R  V  Q  R  P  T  P  R  G  T  V  M  D  I  I  S  R  R  D  Q  R  G  S  G  Q  V  G  T  Y  G  L  N  T

      9410       9420       9430       9440       9450       9460       9470       9480       9490       9500
CTTTCACCAATATGGAAGCCCAACTAATCAGACAGATGGAGGGAGAAGGAGTCTTTAAAAGCATTCAGCACCTAACAATCACAGAAGAAATCGCTGTGCA
  F  T  N  M  E  A  Q  L  I  R  Q  M  E  G  E  G  V  F  K  S  I  Q  H  L  T  I  T  E  E  I  A  V  Q

      9510       9520       9530       9540       9550       9560       9570       9580       9590       9600
AAACTGGTTAGCAAGAGTGGGGCGCGAAAGGTTATCAAGAATGGCCATCAGTGGAGATGATTGTGTTGTGAAACCTTTAGATGACAGGTTCGCAAGCGCT
 N  W  L  A  R  V  G  R  E  R  L  S  R  M  A  I  S  G  D  D  C  V  V  K  P  L  D  D  R  F  A  S  A

      9610       9620       9630       9640       9650       9660       9670       9680       9690       9700
TTAACAGCTCTAAATGACATGGGAAAGATTAGGAAAGACATACAACAATGGGAACCTTCAAGAGGATGGAATGATTGGACACAAGTGCCCTTCTGTTCAC
L  T  A  L  N  D  M  G  K  I  R  K  D  I  Q  Q  W  E  P  S  R  G  W  N  D  W  T  Q  V  P  F  C  S  H

      9710       9720       9730       9740       9750       9760       9770       9780       9790       9800
ACCATTTCCATGAGTTAATCATGAAAGACGGTCGCGTACTCGTTGTTCCATGTAGAAACCAAGATGAACTGATTGGCAGAGCCCGAATCTCCCAAGGAGC
  H  F  H  E  L  I  M  K  D  G  R  V  L  V  V  P  C  R  N  Q  D  E  L  I  G  R  A  R  I  S  Q  G  A

      9810       9820       9830       9840       9850       9860       9870       9880       9890       9900
AGGGTGGTCTTTGCGGGAGACGGCCTGTTTGGGGAAGTCTTACGCCCAAATGTGGAGCTTGATGTACTTCCACAGACGCGACCTCAGGCTGGCGGCAAAT
 G  W  S  L  R  E  T  A  C  L  G  K  S  Y  A  Q  M  W  S  L  M  Y  F  H  R  R  D  L  R  L  A  A  N

      9910       9920       9930       9940       9950       9960       9970       9980       9990      10000
GCTATTTGCTCGGCAGTACCATCACATTGGGTTCCAACAAGTCGAACAACCTGGTCCATACATGCTAAACATGAATGGATGACAACGGAAGACATGCTGA
A  I  C  S  A  V  P  S  H  W  V  P  T  S  R  T  T  W  S  I  H  A  K  H  E  W  M  T  T  E  D  M  L  T
```

FIGURE 14K

```
     10010      10020      10030      10040      10050      10060      10070      10080      10090      10100
CAGTCTGGAACAGGGTGTGGATTCAAGAAAACCCATGGATGGAAGACAAAACTCCAGTGGAATCATGGGAGGAAATCCCATACTTGGGGAAAAGAGAAGA
  V  W  N  R  V  W  I  Q  E  N  P  W  M  E  D  K  T  P  V  E  S  W  E  E  I  P  Y  L  G  K  R  E  D

                          HpaI
     10110      10120      10130      10140      10150      10160      10170      10180      10190      10200
CCAATGGTGCGGCTCATTGATTGGGTTAACAAGCAGGGCCACCTGGGCAAAGAACATCCAAGCAGCAATAAATCAAGTTAGATCCCTTATAGGCAATGAA
  Q  W  C  G  S  L  I  G  L  T  S  R  A  T  W  A  K  N  I  Q  A  A  I  N  Q  V  R  S  L  I  G  N  E

                                                                          3'-NC
     10210      10220      10230      10240      10250      10260      10280      10290      10300
GAATACACAGATTACATGCCATCCATGAAAAGATTCAGAAGAGAAGAGGAAGAAGCAGGAGTTCTGTGGTAGAAAGCAAAACTAACATGAAACAAGGCTA
E  Y  T  D  Y  M  P  S  M  K  R  F  R  R  E  E  E  E  A  G  V  L  W

     10310      10320      10330      10340      10350      10360      10370      10380      10390      10400
GAAGTCAGGTCGGATTAAGCCATAGTACGGAAAAAACTATGCTACCTGTGAGCCCCGTCCAAGGACGTTAAAAGAAGTCAGGCCATCATAAATGCCATAG

     10410      10420      10430      10440      10450      10460      10470      10480      10490      10500
CTTGAGTAAACTATGCAGCCTGTAGCTCCACCTGAGAAGGTGTAAAAAATCCGGGAGGCCACAAACCATGGAAGCTGTACGCATGGCGTAGTGGACTAGC

     10510      10520      10530      10540      10550      10560      10570      10580      10590      10600
GGTTAGAGAGGACCCCTCCCTTACAAATCGCAGCAACAATGGGGGCCCAAGGCGAGATGAAGCTGTAGTCTCGCTGGAAGGACTAGAGGTTAGAGGAGAC

     10610      10620      10630      10640      10650      10660      10670      10680      10690      10700
CCCCCCGAAACAAAAAACAGCATATTGACGCTGGGAAAGACCAGAGATCCTGCTGTCTCCTCAGCATCATTCCAGGCACAGAACGCCAGAAAATGGAATG

     10710      10720
GTGCTGTTGAATCAACAGGTTCT=3' Terminus of the Genomic mRNA

                    CTAGAGCATGCTAGGTGATAA = Amplimer cD2-10687.Xba
                    CTAGATAGGTGATAA = Amplimer cD-10687.X2
                    XbaI
```

FIGURE 14L

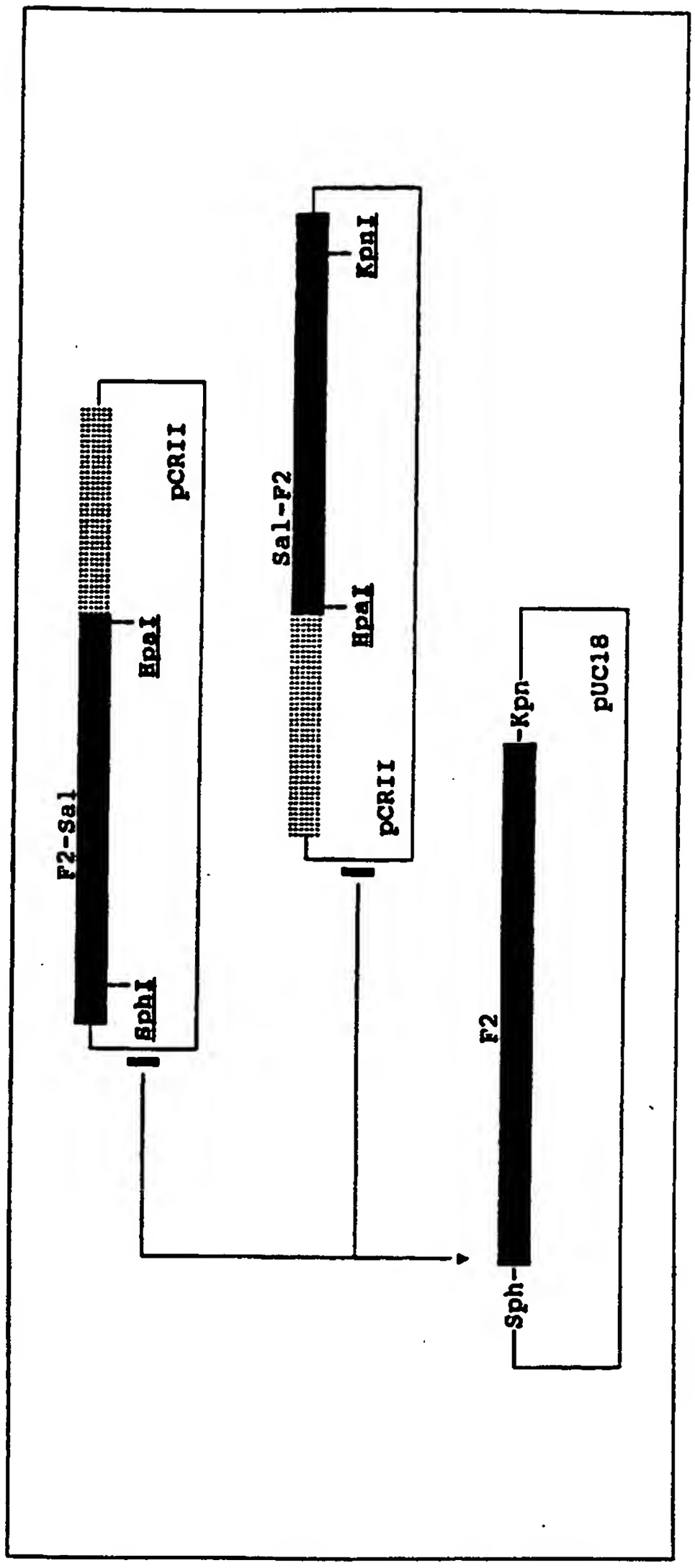


FIGURE 15

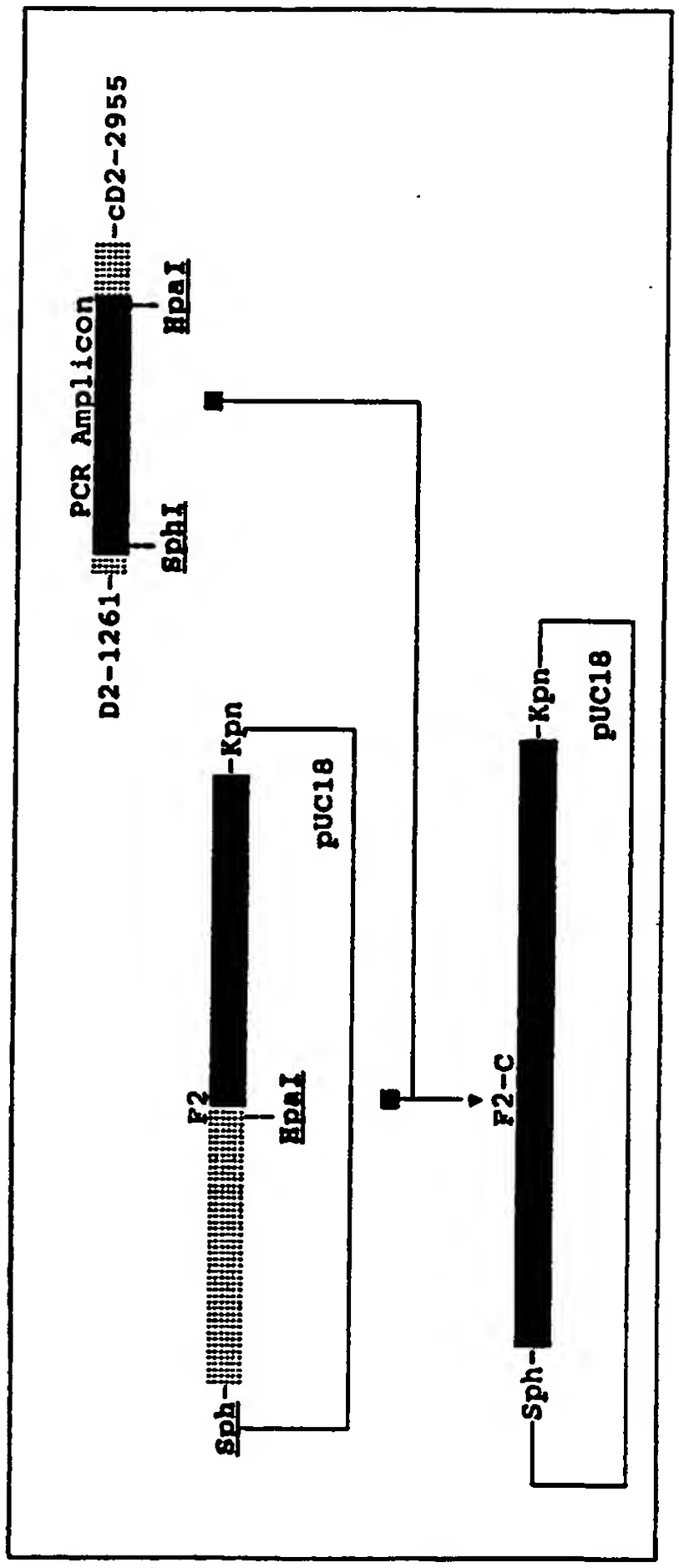

D2-1261
SphI
PCR Amplicon
HpaI
-cD2-2955
pUC18
Kpn
F2
HpaI
Sph-
F2-C
Kpn
pUC18
Sph-

FIGURE 16

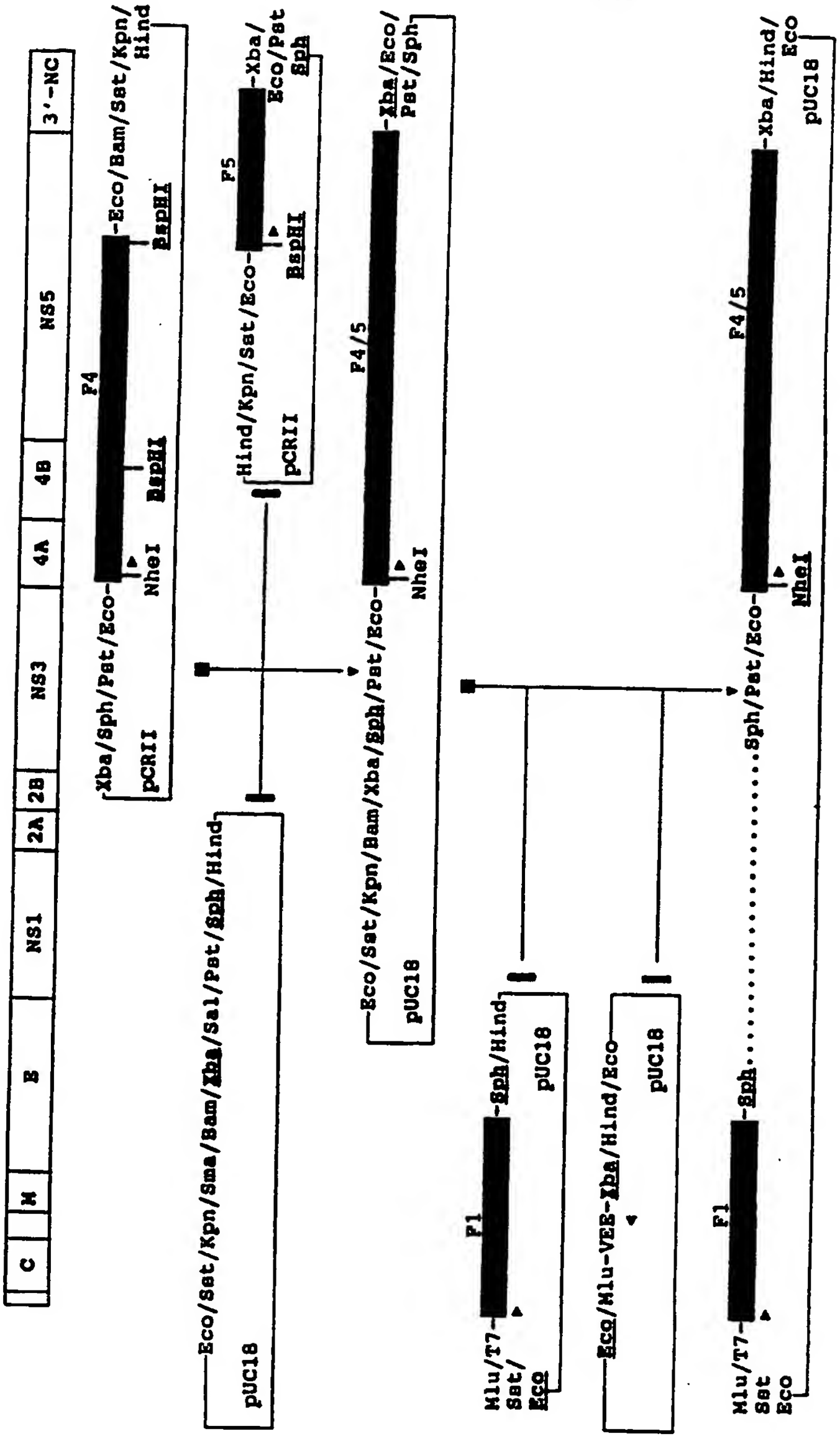

C
M
E
NS1
2A
2B
NS3
4A
4B
3'-NC
NS5
F4
-Eco/Bam/Sst/Kpn/
Hind
BspHI
Xba/Sph/Pst/Eco-
pCRII
NheI
BspHI
F5
-Xba/
Eco/Pst
Sph
Hind/Kpn/Sst/Eco-
pCRII
BspHI
F4/5
-Xba/Eco/
Pst/Sph
Eco/Sst/Kpn/Bam/Xba/Sph/Pst/Eco-
pUC18
NheI
F4/5
-Xba/Hind/
Eco
pUC18
Eco/Sst/Kpn/Sma/Bam/Xba/Sal/Pst/Sph/Hind
pUC18
F1
Mlu/T7-
Sst/
Eco
-Sph/Hind
pUC18
Eco/Mlu-VEE-Xba/Hind/Eco
pUC18
F1
Mlu/T7-
Sst
Eco
-Sph.....Sph/Pst/Eco-
NheI

FIGURE 17A

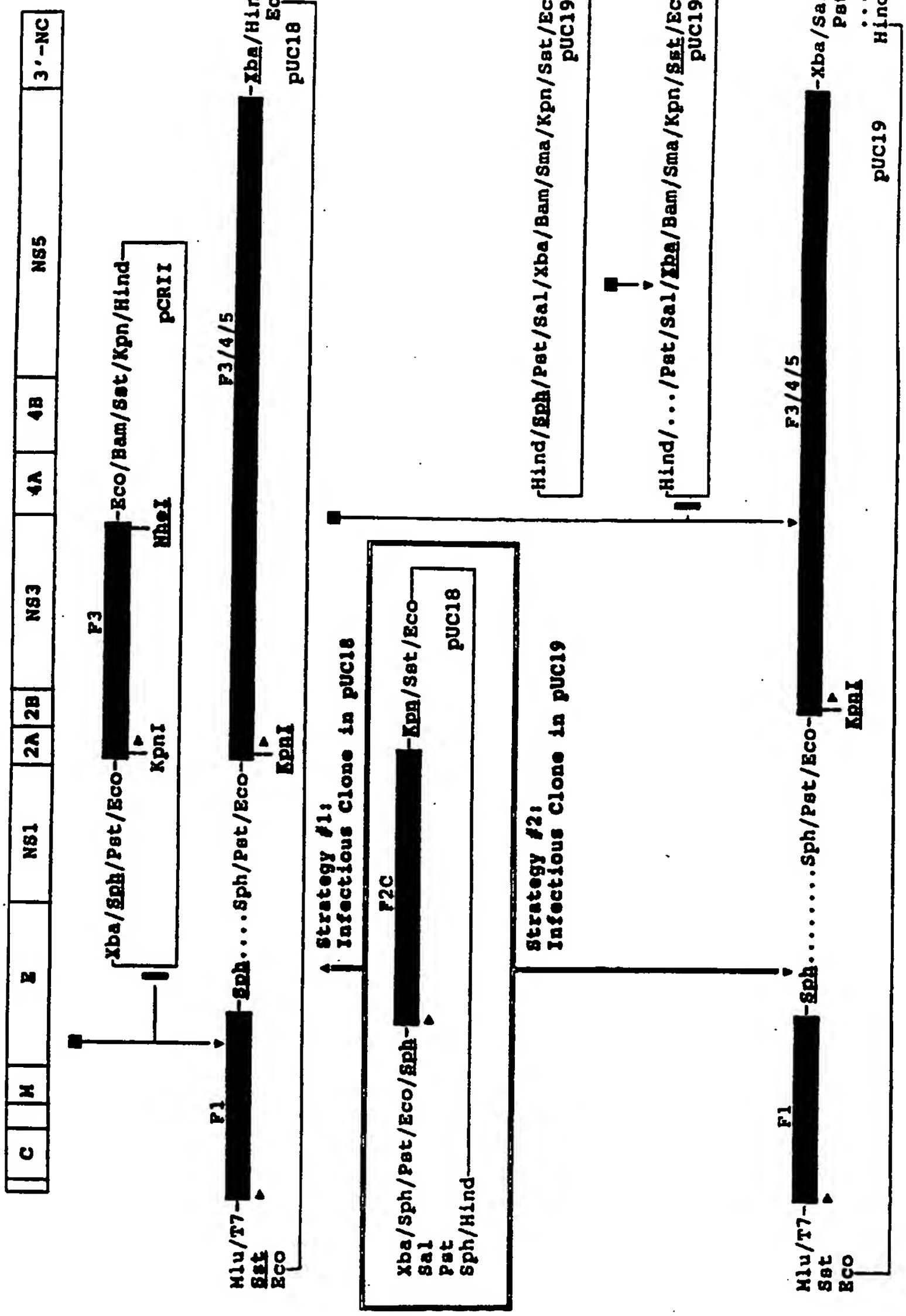

C
M
E
NS1
2A
2B
NS3
4A
4B
NS5
3'-NC
F3/4/5
-Xba/Hind/
Eco
pUC18
pCRII
Xba/Sph/Pst/Eco-
Eco/Bam/Sac/Kpn/Hind
F3
Eco/Bam
NheI
KpnI
-Sph....Sph/Pst/Eco-
KpnI
Strategy #1:
Infectious Clone in pUC18
F2C
Kpn/Sac/Eco
pUC18
P1
Mlu/T7-
Sac
Eco
Strategy #2:
Infectious Clone in pUC19
Xba/Sph/Pst/Eco/Sph-
Sal
Pst
Sph/Hind
Hind/Sph/Pst/Sal/Xba/Bam/Sma/Kpn/Sac/Eco
pUC19
Hind/.../Pst/Sal/Xba/Bam/Sma/Kpn/Sac/Eco
pUC19
F3/4/5
-Xba/Sal/
Pst/
...
Hind
pUC19
-Sph.....Sph/Pst/Eco-
KpnI
F1
Mlu/T7-
Sac
Eco

FIGURE 17B

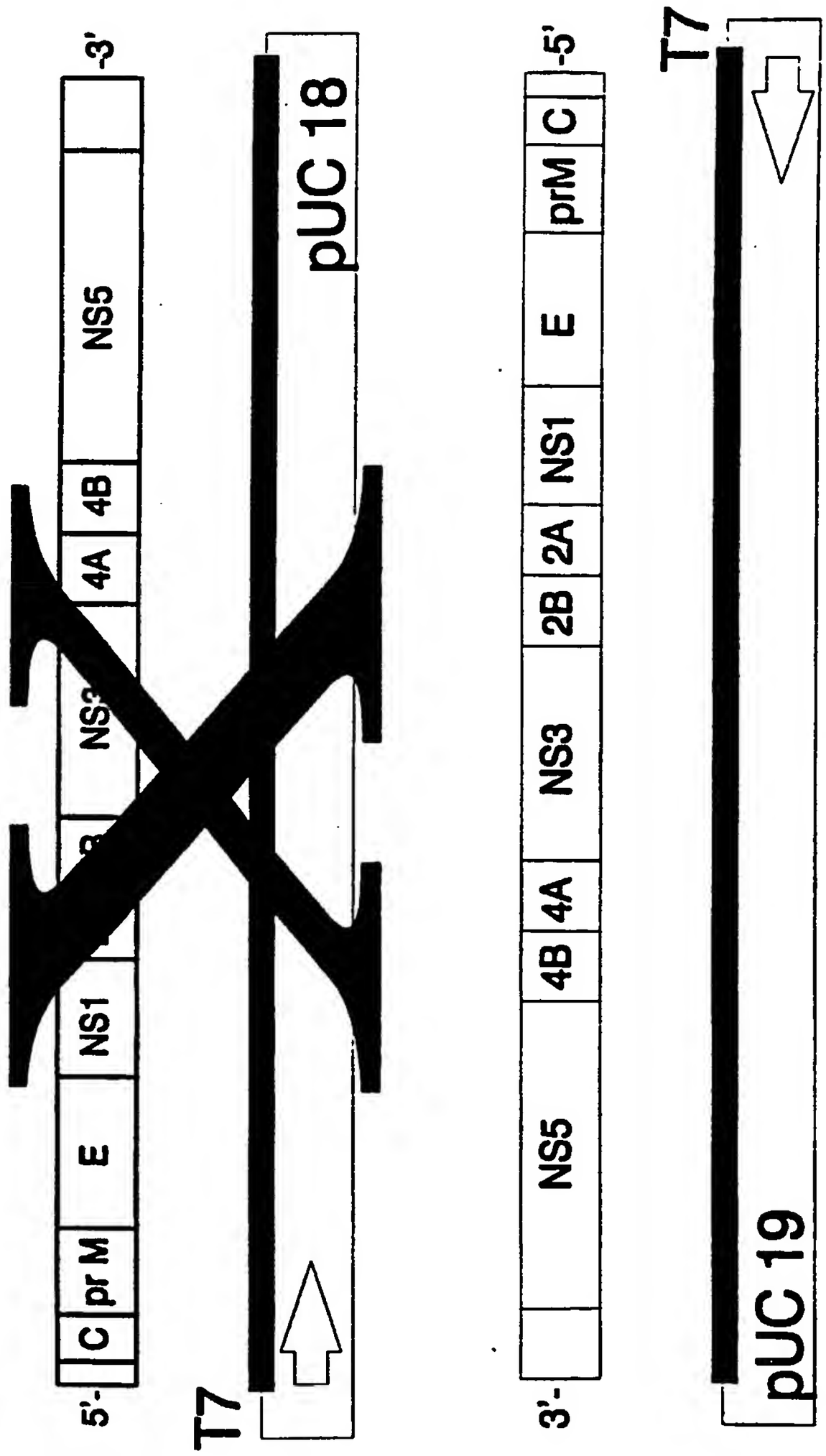


FIGURE 18

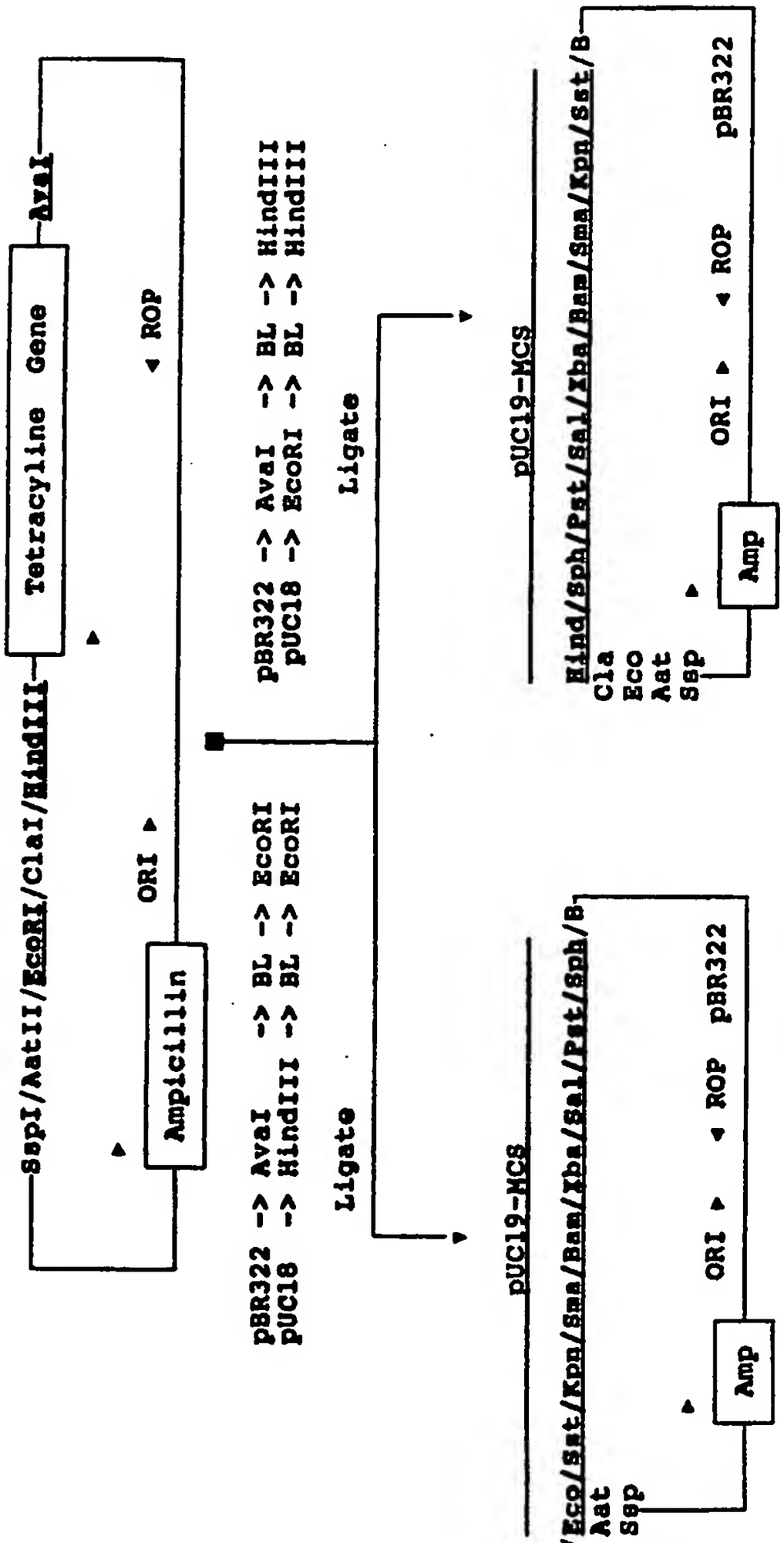

SspI/AatII/EcoRI/ClaI/HindIII
Tetracycline Gene
AvaI
ROP
ORI
Ampicillin
pBR322 -> AvaI -> BL -> EcoRI
pUC18 -> HindIII -> BL -> EcoRI
Ligate
pUC19-MCS
Hind/Sph/Pst/Sal/Xba/Bam/Sma/Kpn/Sst/B
Cla
Eco
Aat
Sap
Amp
pBR322
pBRUC-139
pBRUC-138

FIGURE 19

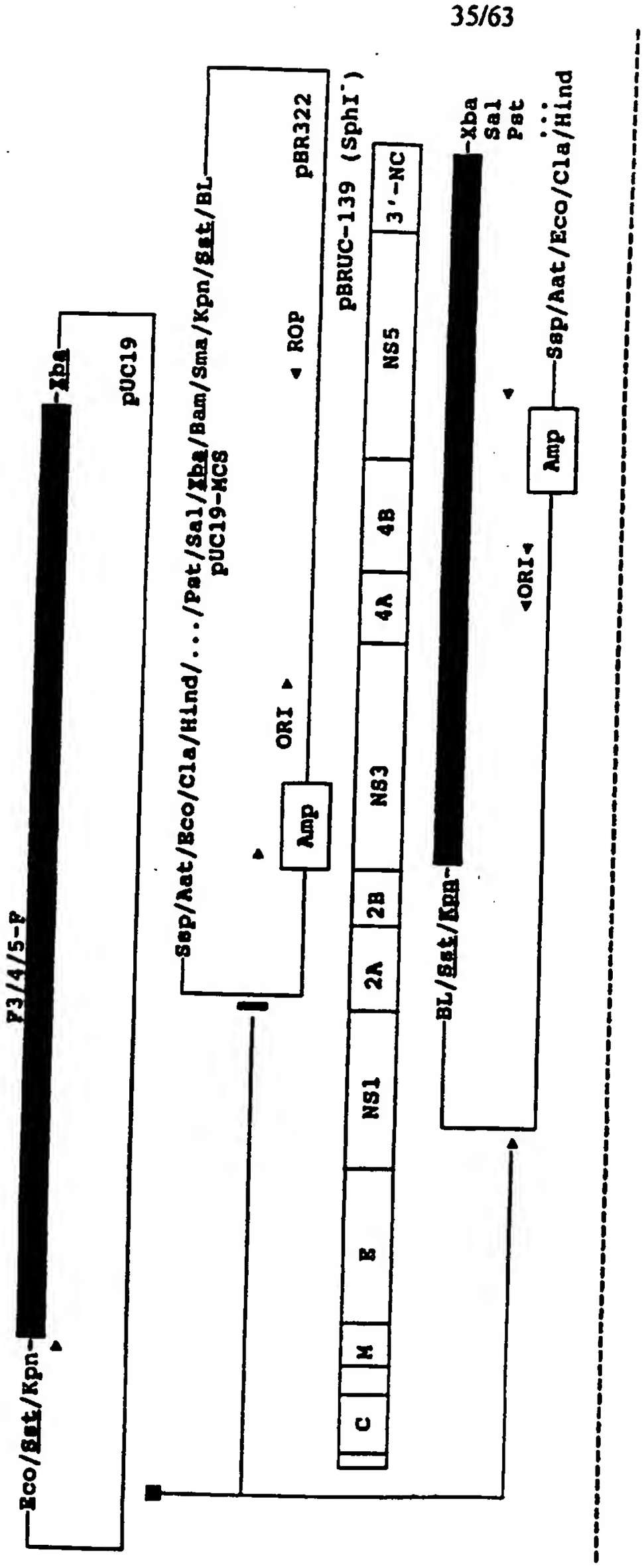

P3/4/5-P
Xba
pUC19
Sma/Kpn/Sst/BL
pBR322
ROP
pBRUC-139 (SphI⁻)
NS5
3'-NC
Xba
Sal
Pst
Sap/Aat/Eco/Cla/Hind...
Sap/Aat/Eco/Cla/Hind/.../Pst/Sal/Xba/Bam/S
pUC19-MCS
ORI
Amp
NS1
2A
2B
NS3
4A
4B
BL/Sst/Kpn
ORI
Amp
C
M
E
Eco/Sst/Kpn

FIGURE 20A

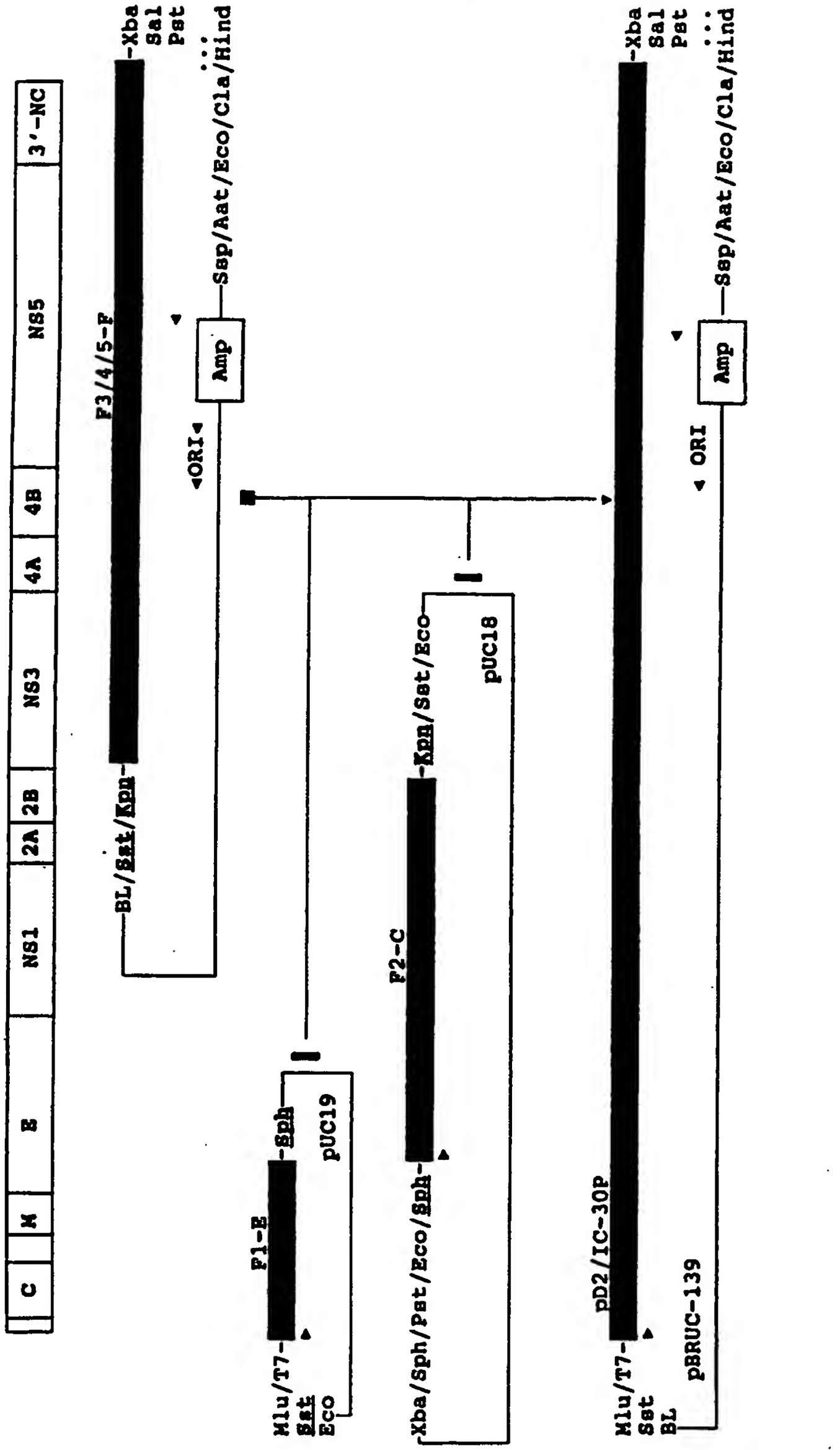

C
M
E
NS1
2A
2B
NS3
4A
4B
NS5
3'-NC
F3/4/5-F
-Xba
Sal
Pst
...Sap/Aat/Eco/Cla/Hind
Amp
ORI
BL/Sst/Kpn
F2-C
-Kpn/Sst/Eco
pUC18
F1-E
-Eph
pUC19
Mlu/T7-
Sst
Eco
Xba/Sph/Pst/Eco/Eph
pD2/IC-30P
Mlu/T7-
Sst
BL
pBRUC-139

FIGURE 20B

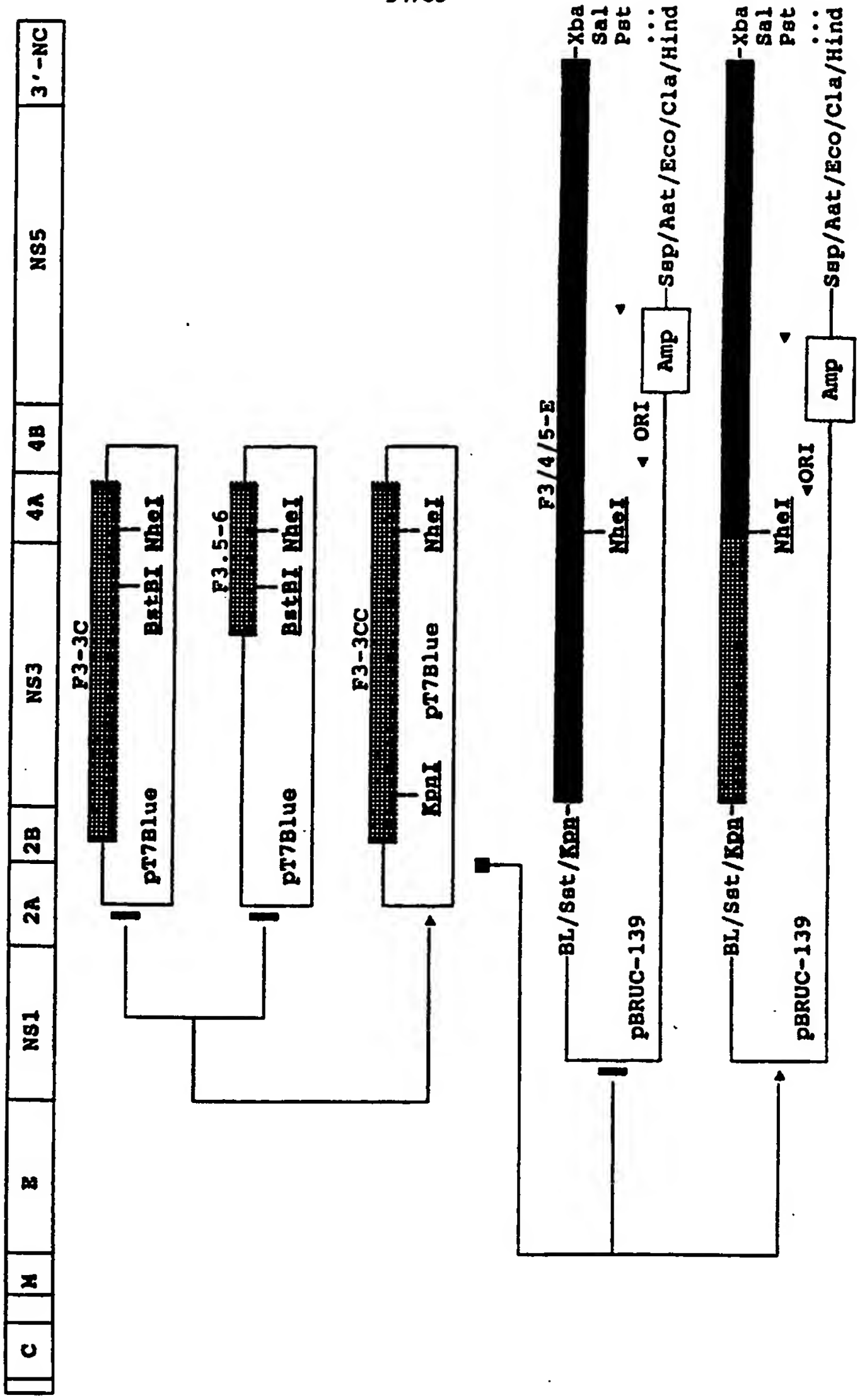

C
M
E
NS1
2A
2B
NS3
4A
4B
NS5
3'-NC
F3-3C
BstBI NheI
pT7Blue
F3.5-6
BstBI NheI
pT7Blue
F3-3CC
KpnI pT7Blue NheI
F3/4/5-E
BL/Sst/Kpn
pBRUC-139
NheI
ORI
Amp
Sap/Aat/Eco/Cla/Hind
Xba
Sal
Pst
...

FIGURE 21A

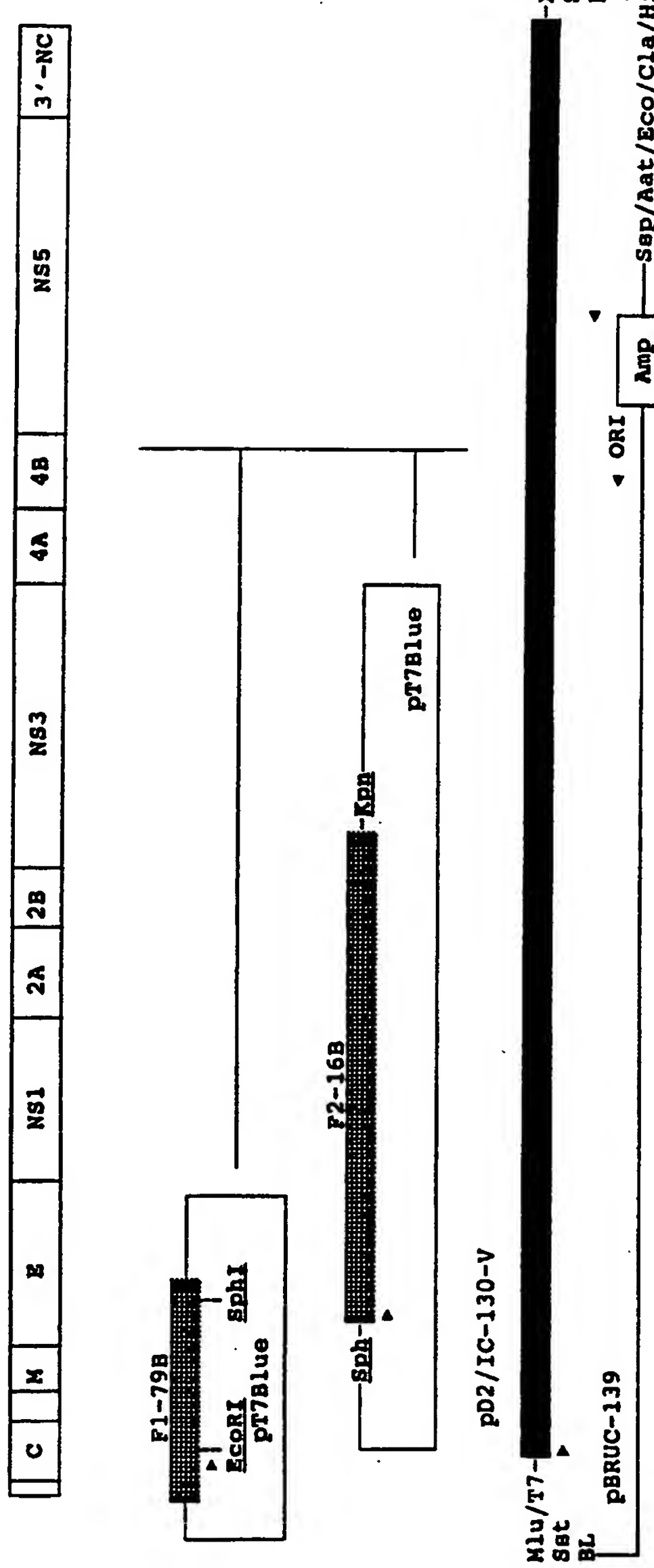

C
M
E
NS1
2A
2B
NS3
4A
4B
NS5
3'-NC
F1-79B
EcoRI
SphI
pT7Blue
F2-16B
Kpn
Sph
pT7Blue
pD2/IC-130-V
Mlu/T7-
Sst
BL
pBRUC-139
Xba
Sal
Pst
...
Sap/Aat/Eco/Cla/Hind
Amp
ORI

FIGURE 21B

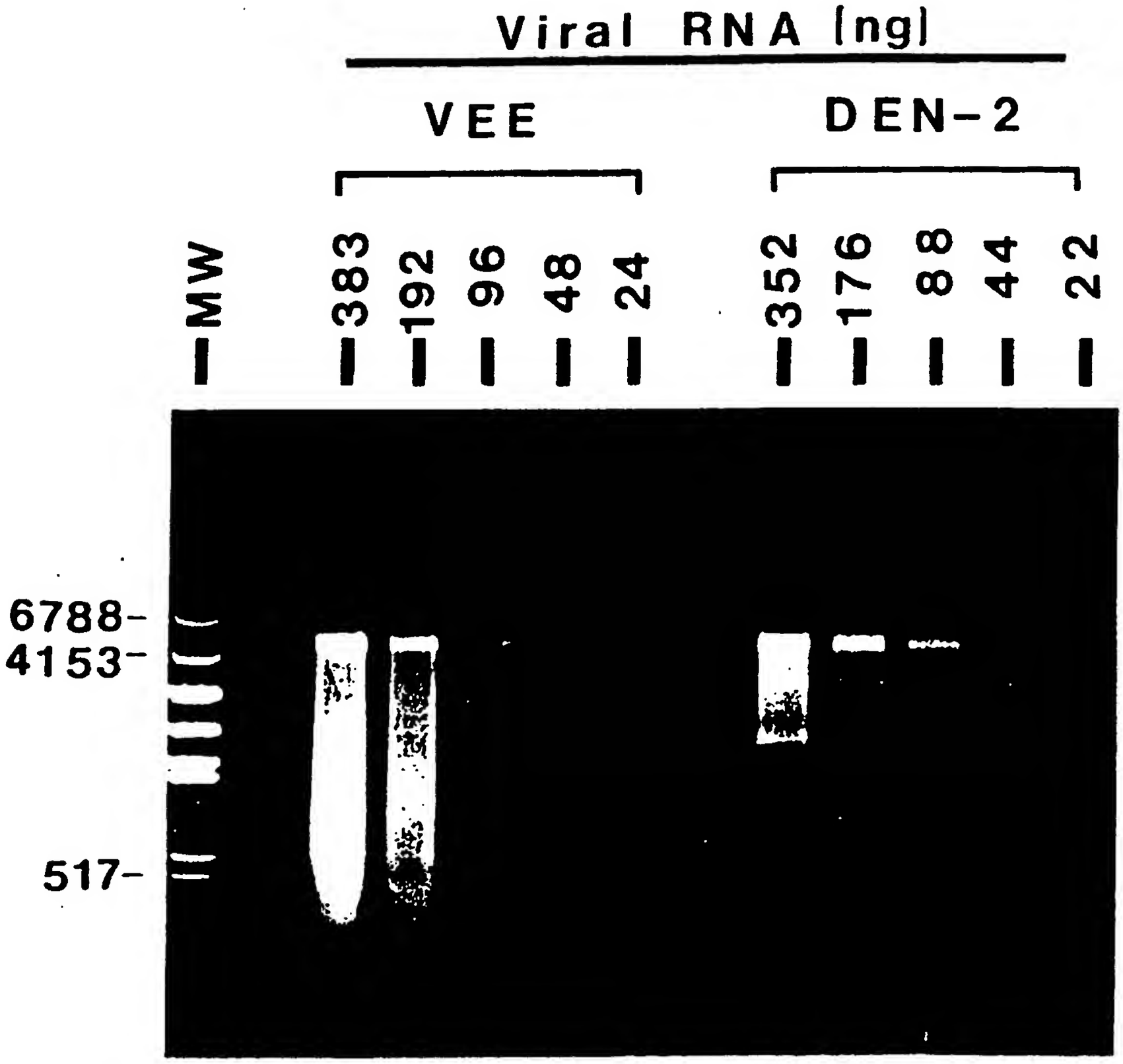

Viral RNA (ng)
VEE
DEN-2
MW
383
192
96
48
24
352
176
88
44
22
6788-
4153-
517-

FIGURE 22

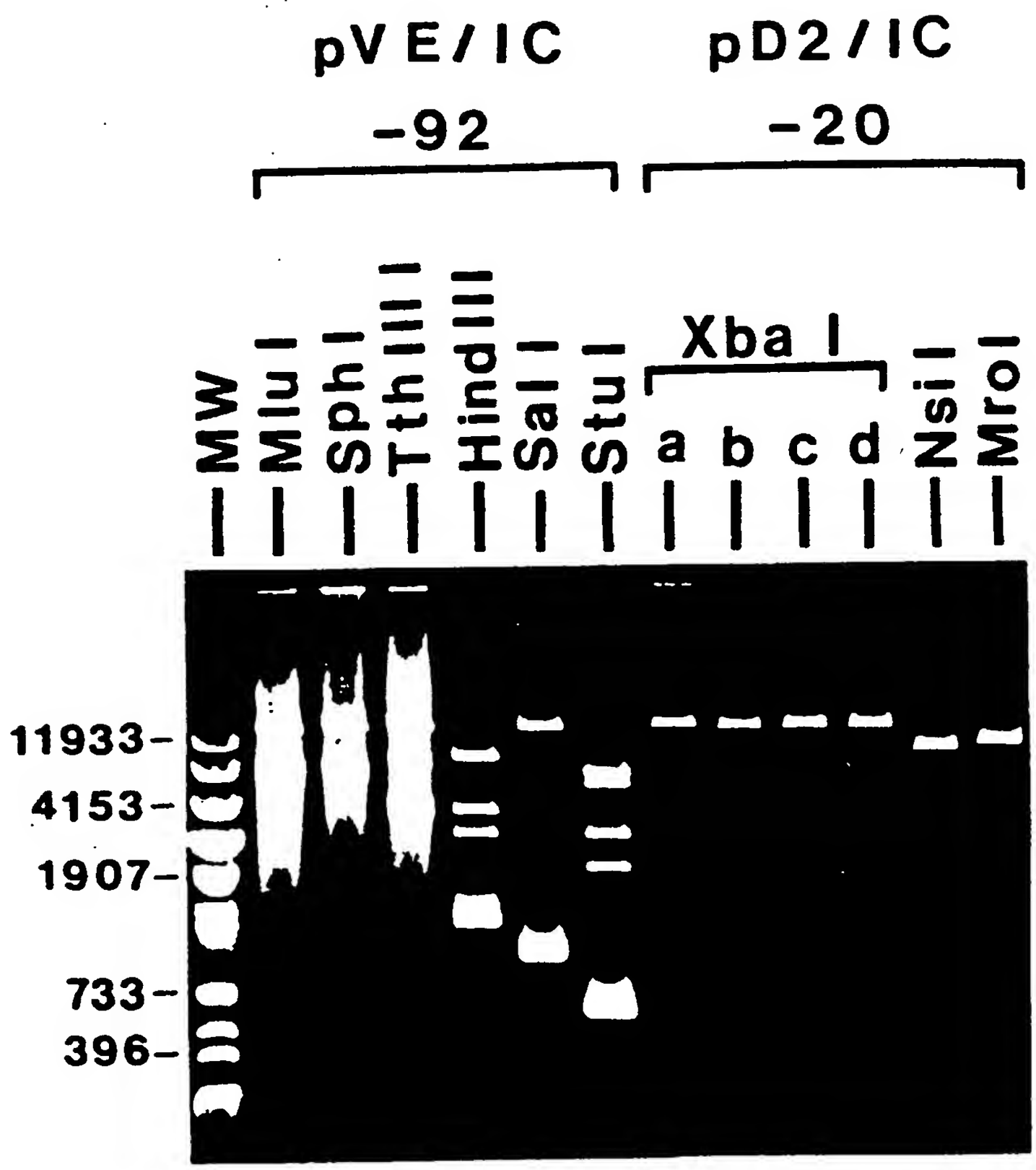

pVE/IC
-92
pD2/IC
-20
MW
MluI
SphI
TthIIII
HindIII
SalI
StuI
Xba I
a
b
c
d
NsiI
MroI
11933-
4153-
1907-
733-
396-

FIGURE 23

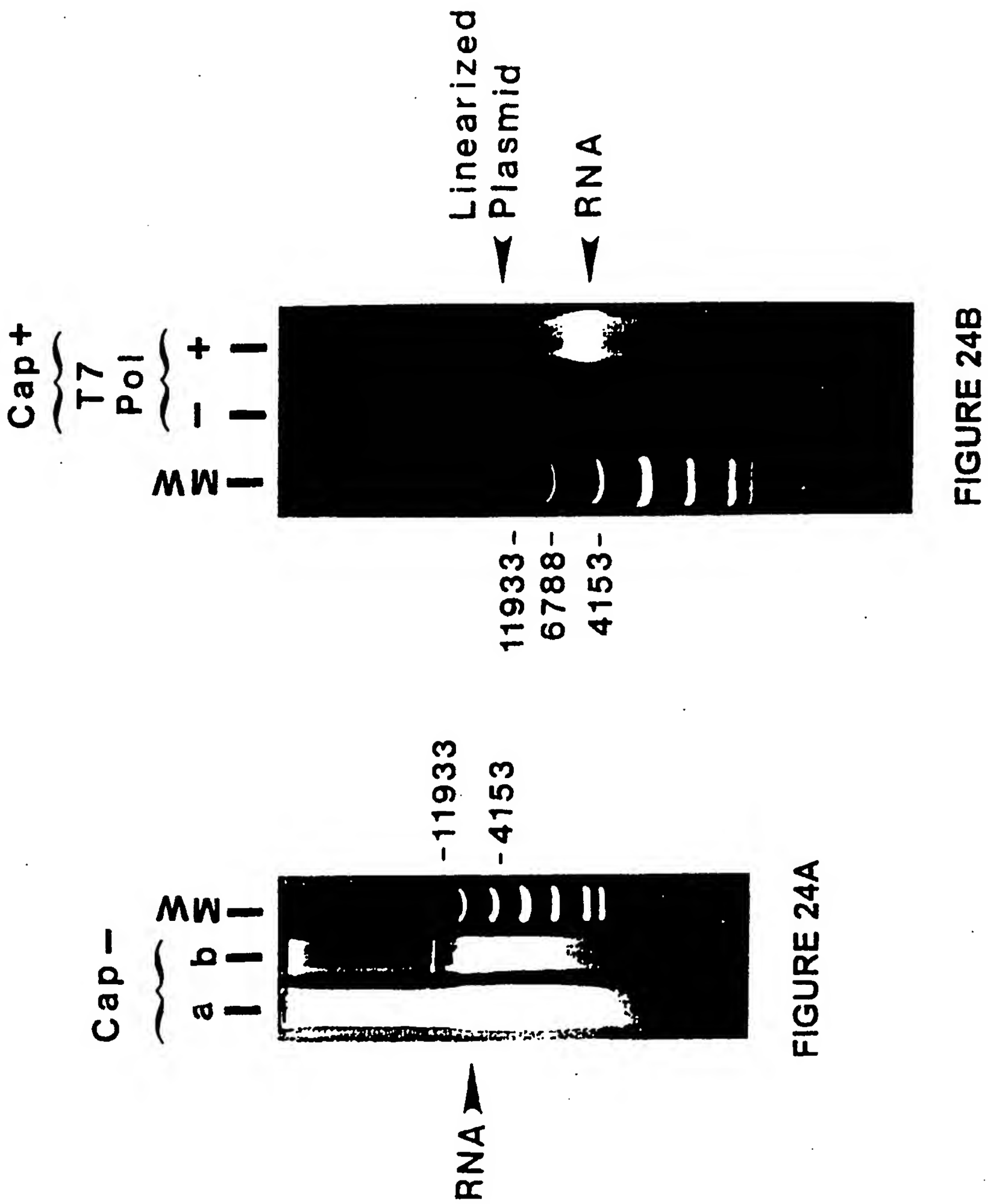


FIGURE 24B

FIGURE 24A

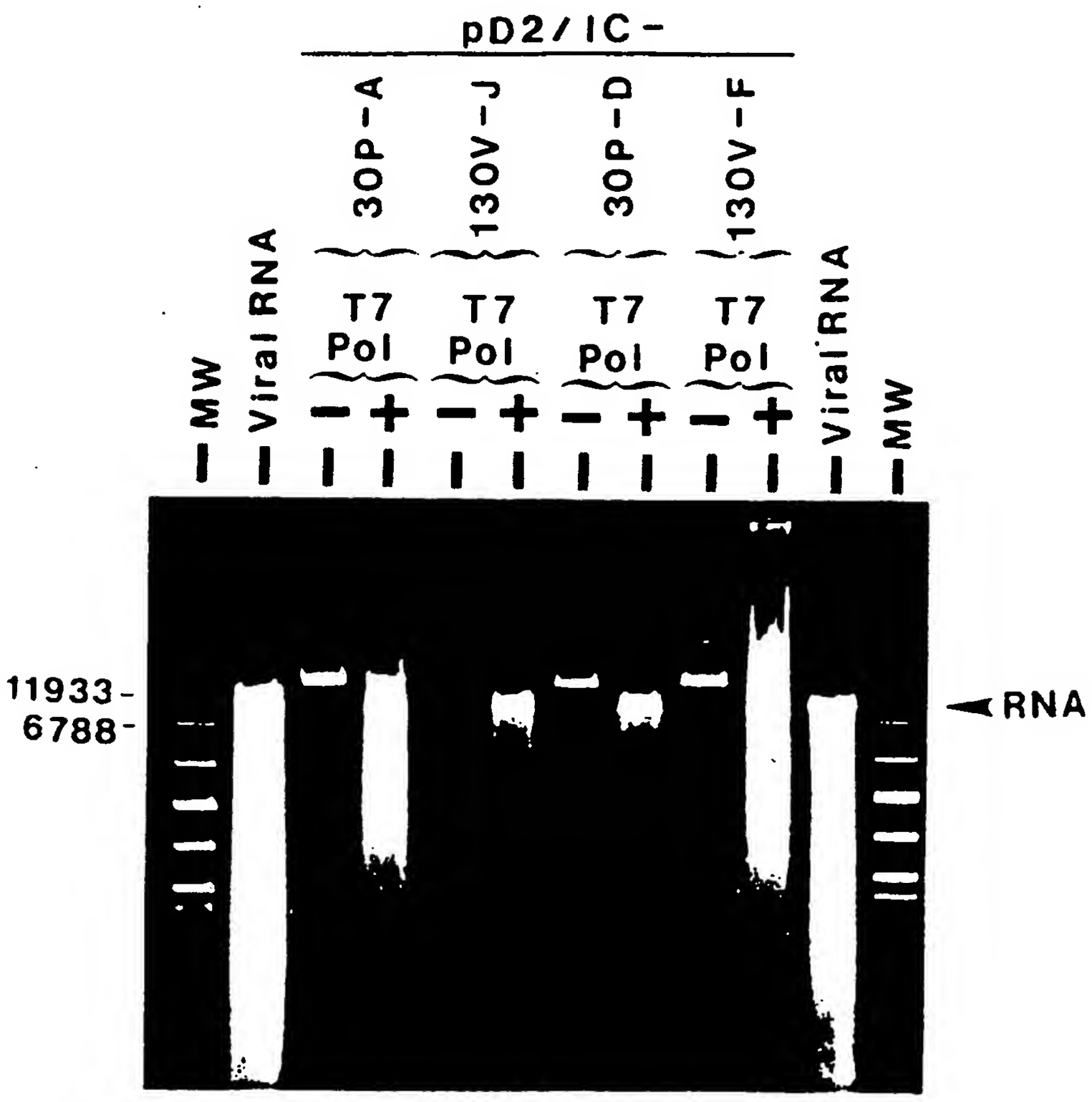

pD2/IC-
P-A
V-J
P-D
V-F
30
130
30
130
MW
Viral RNA
T7 Pol
T7 Pol
T7 Pol
T7 Pol
Viral RNA
MW
11933-
6788-
RNA

FIGURE 25

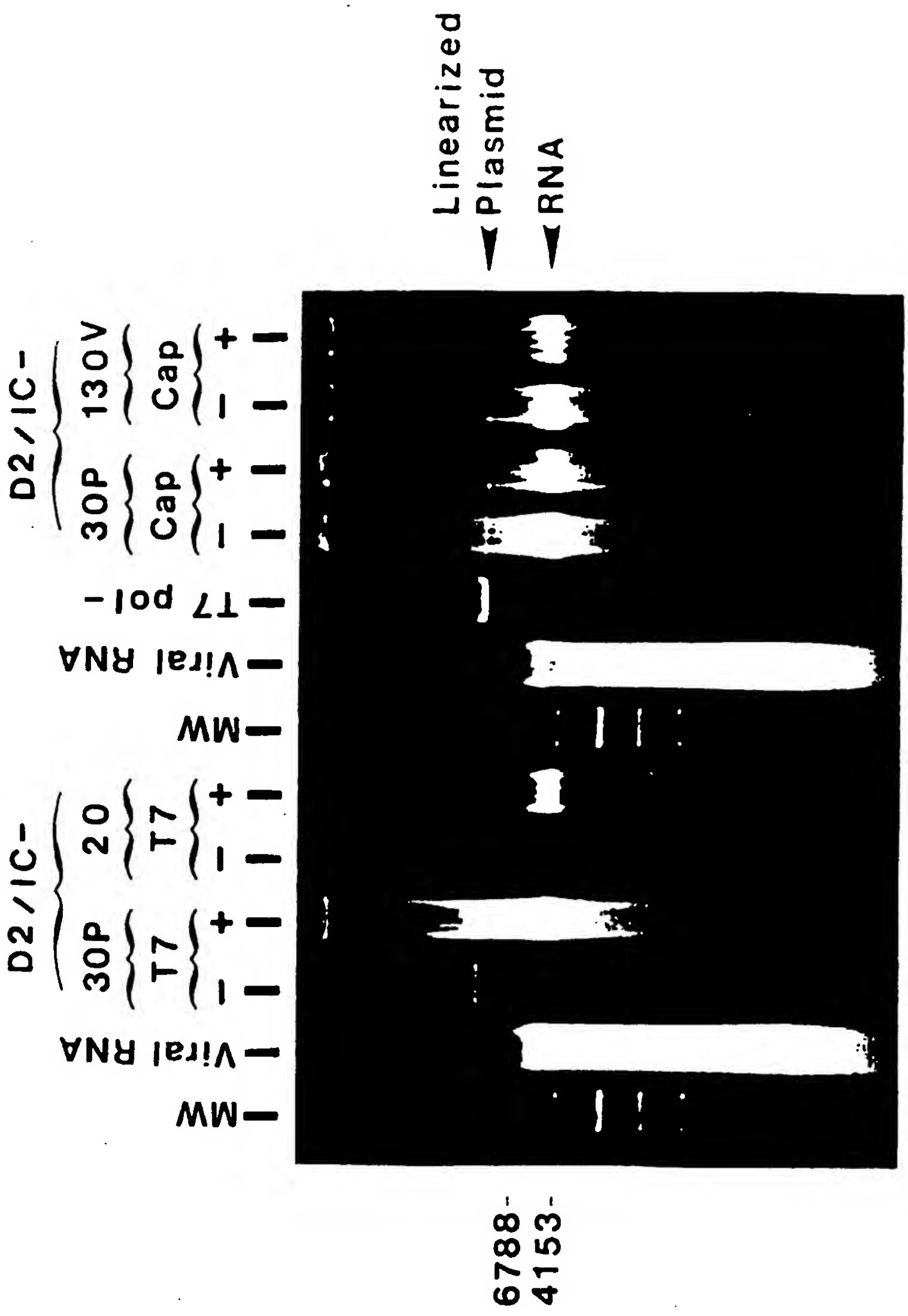

Linearized Plasmid
RNA
D2/IC-
30P
Cap
+
-
T7 pol-
Viral RNA
MW
D2/IC-
30P
130V
Cap
Cap
+
-
T7 pol-
Viral RNA
MW
D2/IC-
30P
20
T7
+
-
Viral RNA
MW
6788-
4153-

FIGURE 26

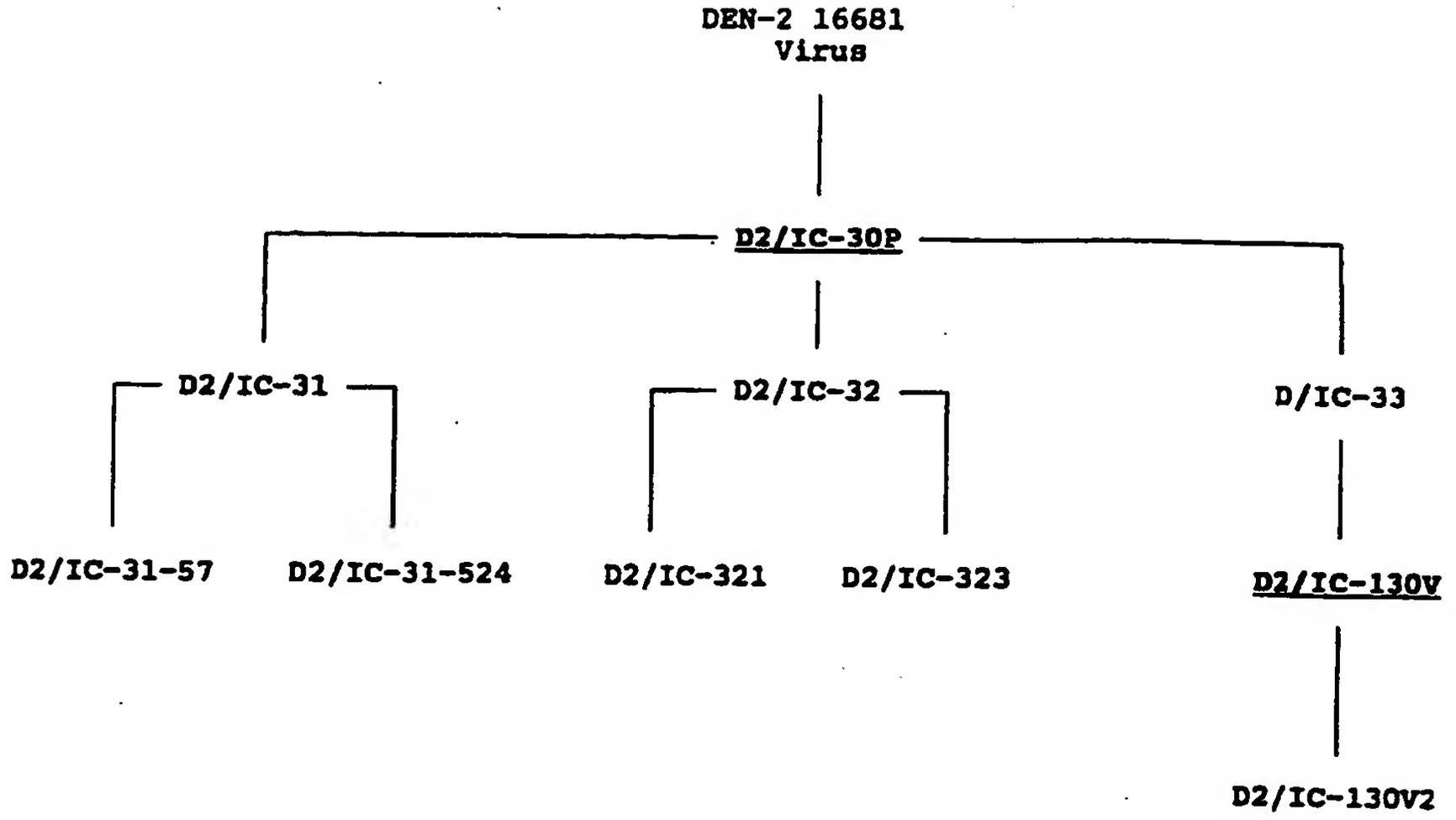

DEN-2 16681
Virus
D2/IC-30P
D2/IC-31
D2/IC-32
D/IC-33
D2/IC-31-57
D2/IC-31-524
D2/IC-321
D2/IC-323
D2/IC-130V
D2/IC-130V2

FIGURE 27

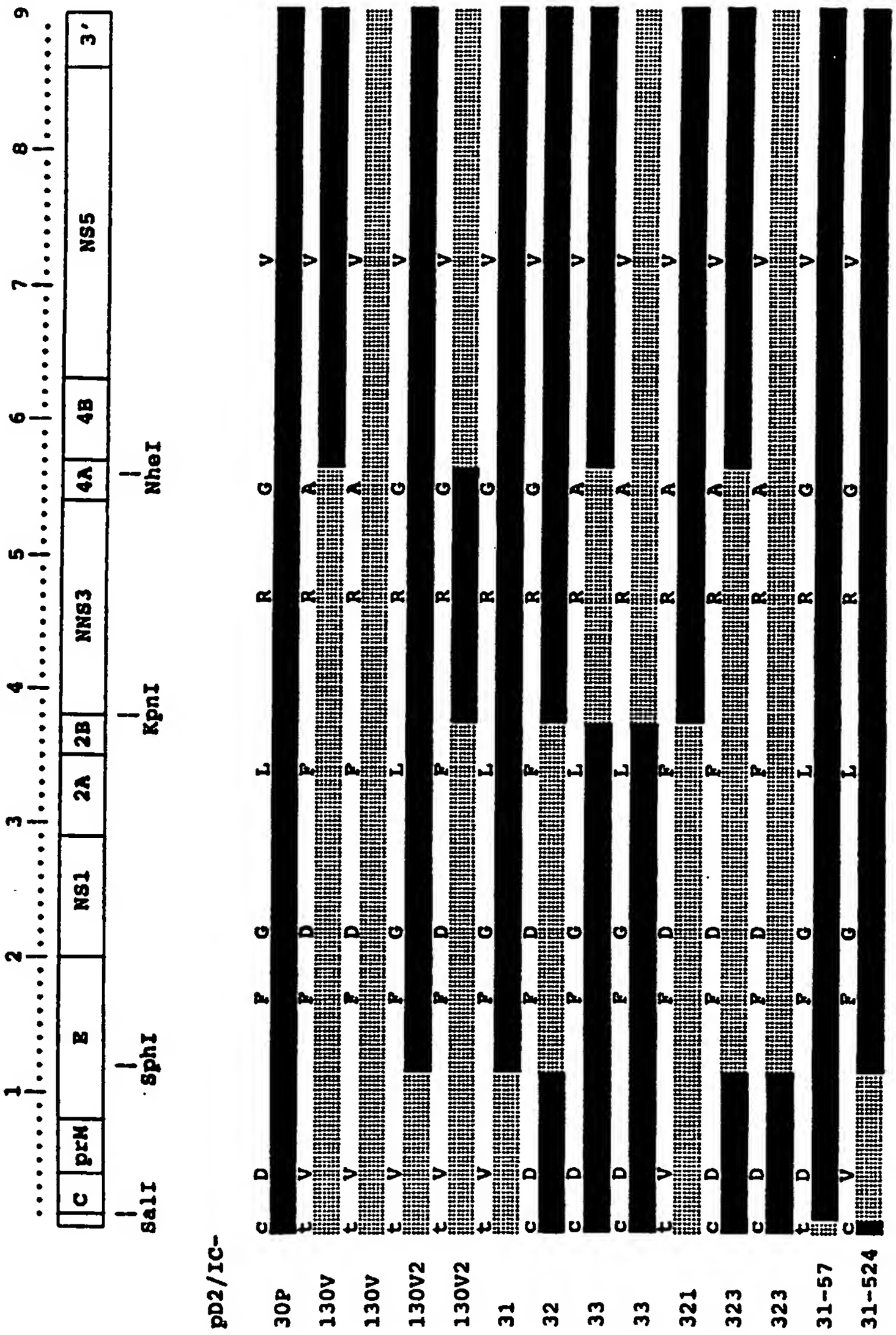

1
2
3
4
5
6
7
8
9
C
prM
E
NS1
2A
2B
NNS3
4A
4B
NS5
3'
SalI
SphI
KpnI
NheI
pD2/IC-
30P
130V
130V
130V2
130V2
31
32
33
33
321
323
323
31-57
31-524

FIGURE 28

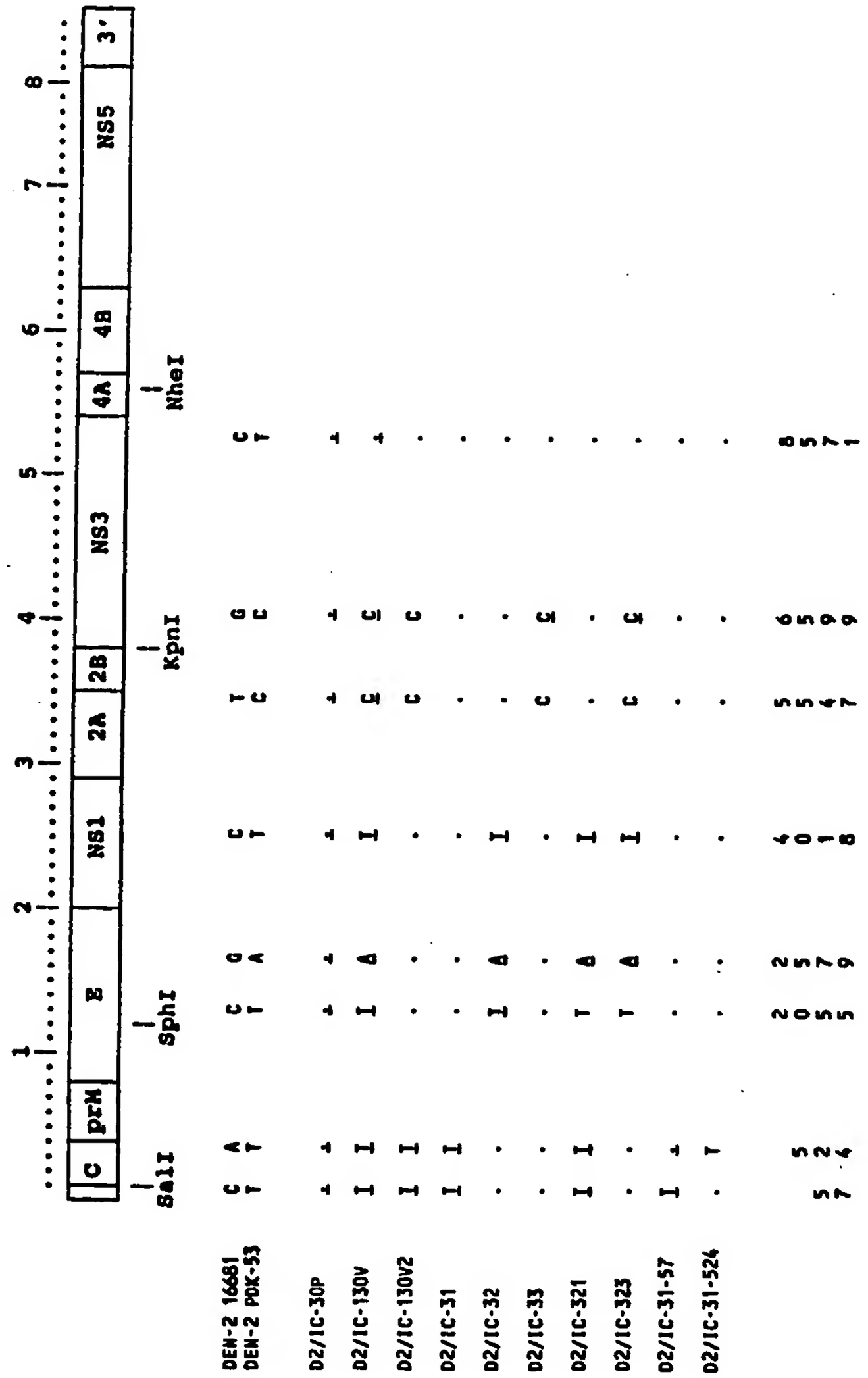

C
prM
E
NS1
2A
2B
NS3
4A
NS5
3'
SalI
SphI
KpnI
NheI
DEN-2 16681
DEN-2 PDK-53
D2/IC-30P
D2/IC-130V
D2/IC-130V2
D2/IC-31
D2/IC-32
D2/IC-33
D2/IC-321
D2/IC-323
D2/IC-31-57
D2/IC-31-524

FIGURE 29

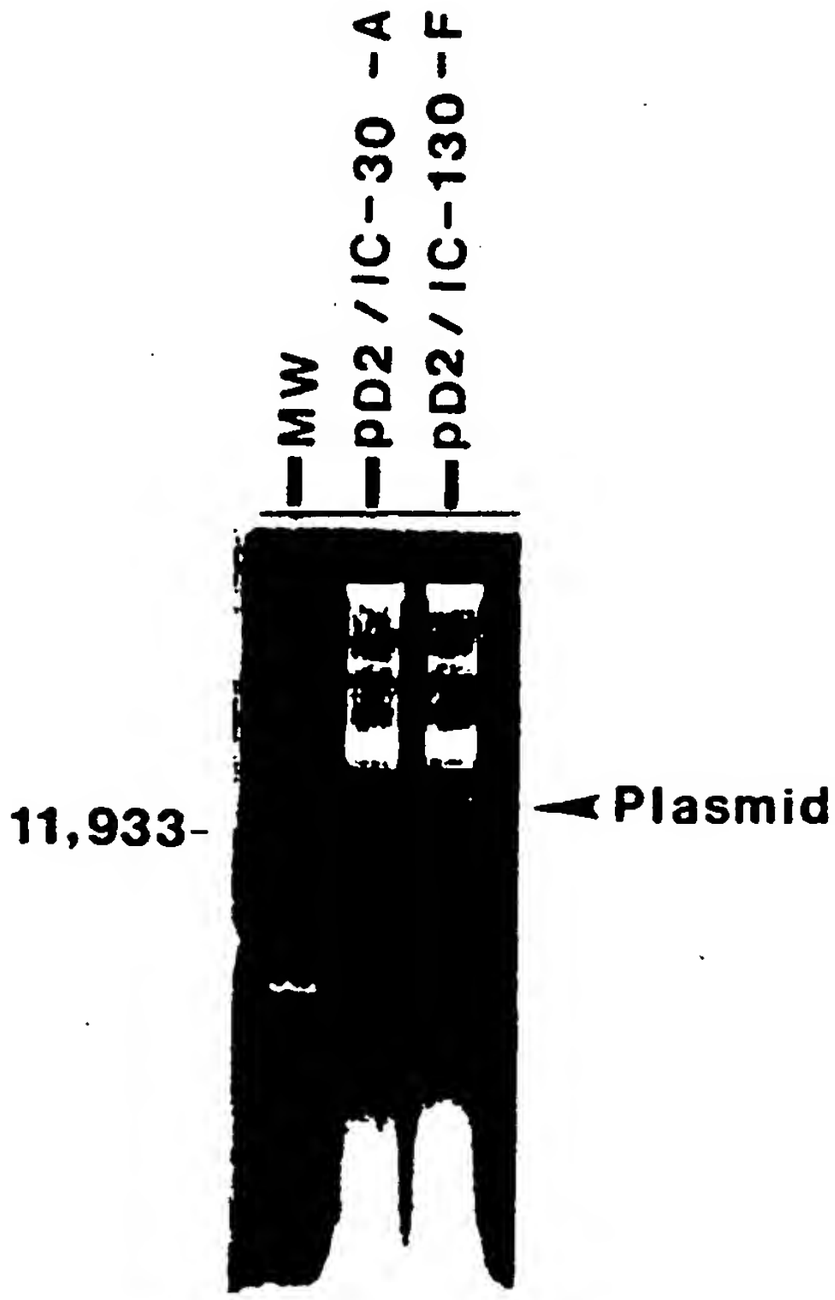


FIGURE 30

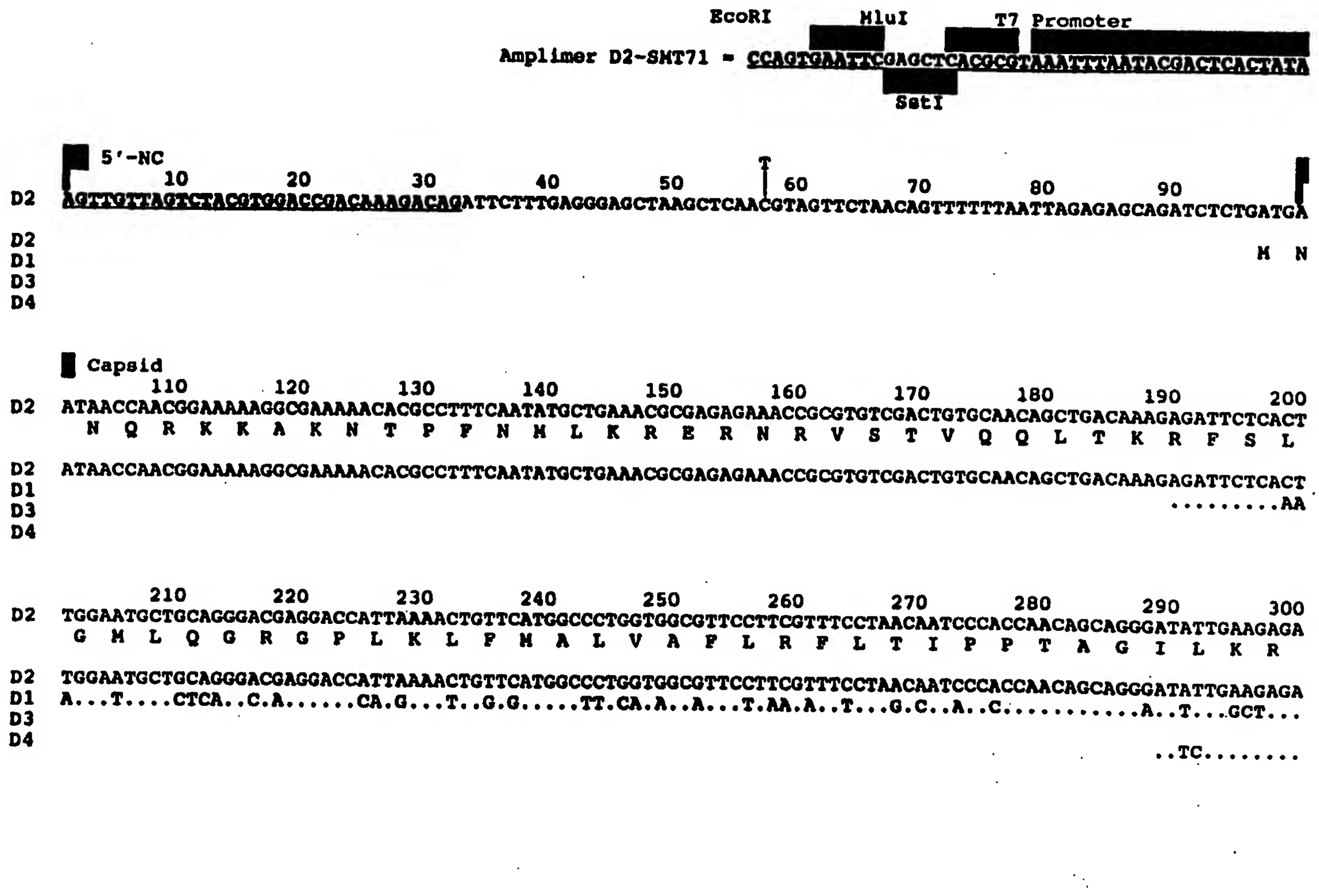

EcoRI
MluI
T7 Promoter
Amplimer D2-SMT71 = CCAGTGAATTCGAGCTCACGCGTAAATTTAATACGACTCACTATA
SstI
5'-NC
Capsid
D2
D1
D3
D4
M N
N Q R K K A K N T P F N M L K R E R N R V S T V Q Q L T K R F S L
G M L Q G R G P L K L F M A L V A F L R F L T I P P T A G I L K R
.....AA
..TC.....

FIGURE 31A

```
          310       320       330       340       350       360       370       380       390       400
D2  TGGGGAACAATTAAAAAATCAAAAGCTATTAATGTTTTGAGAGGGTTCAGGAAAGAGATTGGAAGGATGCTGAACATCTTGAATAGGAGACGCAGATCTG
    W  G  T  I  K  K  S  K  A  I  N  V  L  R  G  F  R  K  E  I  G  R  M  L  N  I  L  N  R  R  R  R  S  A

D2  TGGGGAACAATTAAAAAATCAAAAGCTATTAATGTTTTGAGAGGGTTCAGGAAAGAGATTGGAAGGATGCTGAACATCTTGAATAGGAGACGCAGATCTG
D1  .....CT..T.C..G..GAATGG...G.....A..G..AC.G..T....A..G...A..CTC..AC.....A.....AA....C.....GAAA.....C.
D3                             ..G..C..A.A...C....A...G.....CTC..AC.......G...TA.C..C.AAC.GAAA.AGA.AT
D4  ......CAGT.G..G...AAT..G..C..C.GGA.AC...TT..A........G.....A..CC.C..................CG.....AAA..G..AA

                                                  prM
          410       420       430                 450       460       470       480       490       500
D2  CAGGCATGATCATTATGCTGATTCCAACAGTGATGGCGTTCCATTTAACCACACGTAACGGAGAACCACACATGATCGTCAGCAGACAAGAGAAAGGGAA
      G  M  I  I  M  L  I  P  T  V  M  A  F  H  L  T  T  R  N  G  E  P  H  M  I  V  S  R  Q  E  K  G  K

D2  CAGGCATGATCATTATGCTGATTCCAACAGTGATGGCGTTCCATTTAACCACACGTAACGGAGAACCACACATGATCGTCAGCAGACAAGAGAAAGGGAA
D1  TGAC....C..C........C.G..C....CCC..........C.G..G.....AGGG.....G..G..T.....A..T....AG..G..A.G...A..
D3  .GCT.TGTC....G...A..T.A...G..ACAC.T..T.....C.....TT....AG.T.....G..G.G......T..GG.G.AGA.T..A.G...A..
D4  .GATA.CAT.GC.GTGCT.......C..C..A........T..C..GT.A...A.AG.T..C.....C.T......A..GGCA.A...T..A.GG....G

                    T
          510       520       530       540       550       560       570       580       590       600
D2  AAGTCTTCTGTTTAAAACAGAGGATGGCGTGAACATGTGTACCCTCATGGCCATGGACCTTGGTGAATTGTGTGAAGACACAATCACGTACAAGTGTCCC
     S  L  L  F  K  T  E  D  G  V  N  M  C  T  L  M  A  M  D  L  G  E  L  C  E  D  T  I  T  Y  K  C  P
                    V

D2  AAGTCTTCTGTTTAAAACAGAGGTTGGCGTGAACATGTGTACCCTCATGGCCATGGACCTTGGTGAATTGTGTGAAGACACAATCACGTACAAGTGTCCC
D1  GTCA...T....C..G..CTCT.CA..T..C........C........T..G.....TT.G..A..G........G.....G..G..C.....A..C...
D3  .TCC..A..T..C..G....CCTC...AA.C........C..A.....A..T.....T..G..A..GA.......T.....GG....T.....A..C...
D4  .CC...CT.......G...ACA.AG..GA.C....AA..C..T.....T.........T.G......A.......G.....TG.......T..A..C...
```

FIGURE 31B

```
                610       620       630       640       650       660       670       680       690       700
D2  CTTCTCAGGCAGAATGAGCCAGAAGACATAGACTGTTGGTGCAACTCTACGTCCACGTGGGTAACTTATGGGACGTGTACCACCATGGGAGAACATAGAA
    L  L  R  Q  N  E  P  E  D  I  D  C  W  C  N  S  T  S  T  W  V  T  Y  G  T  C  T  T  M  G  E  H  R  R

D2  CTTCTCAGGCAGAATGAGCCAGAAGACATAGACTGTTGGTGCAACTCTACGTCCACGTGGGTAACTTATGGGACGTGTACCACCATGGGAGAACATAGAA
D1  .GGA...CTG..GCG..A.....T...G.T..............TG.C...GA...A.....G..C.....A.....CT.TCAA.CT..C.....CC..C
D3  .ACA.T.CCG.AGTG.....T........T.....C.........CT...A..G..A.....G........A..A..C.ATCAAGCT.....G......C
D4  T.A..GGTCA.T.CC..A..T........T..T..C........TCTC.....T..C.....C..A........A.AC...CAG.GC.......GG...C

                    M
                710 |     720       730       740       750       760       770       780       790       800
D2  GAGAAAAAAGATCAGTGGCACTCGTTCCACATGTGGGAATGGGACTGGAGACACGAACTGAAACATGGATGTCATCAGAAGGGGCCTGGAAACATGTCCA
      E  K  R  S  V  A  L  V  P  H  V  G  M  G  L  E  T  R  T  E  T  W  M  S  S  E  G  A  W  K  H  V  Q

D2  GAGAAAAAAGATCAGTGGCACTCGTTCCACATGTGGGAATGGGACTGGAGACACGAACTGAAACATGGATGTCATCAGAAGGGGCCTGGAAACATGTCCA
D1  ....C...C.T..C..C...T.G.CC.....C.....GC.T..C..A..A...A..G.C.....G........C..T.....T..T........GA.A..
D3  .C..T..G...........GT.A.C...C.....T..C...........C.....C...C....C........GG.T.....A..T....G...A...G.
D4  ....G..GC.C.....A..TT.AACA......TCA.........T....A...A..G....G.............G........T.....G....CT..

                810       820       830       840       850       860       870       880       890       900
D2  GAGAATTGAAACTTGGATCTTGAGACATCCAGGCTTCACCATGATGGCAGCAATCCTGGCATACACCATAGGAACGACACATTTCCAAAGAGCCCTGATT
     R  I  E  T  W  I  L  R  H  P  G  F  T  M  M  A  A  I  L  A  Y  T  I  G  T  T  H  F  Q  R  A  L  I

D2  GAGAATTGAAACTTGGATCTTGAGACATCCAGGCTTCACCATGATGGCAGCAATCCTGGCATACACCATAGGAACGACACATTTCCAAAGAGCCCTGATT
D1  A.A.G.A..G......GCTC.............A.....GG....A..CCTTT.T..A...C.TG..........AT.CATCAC...G.A..GGA.C...
D3  ..AGG.A..G..A...GC.C.T..G..C.....G..T.....AC.A..CCT.T.T..T..CC.TTA......C..TT.CTTGAC...G.A..TGG.T...
D4  ....G.A..G.GC.....AC.C...A.C.....A...G.GC.CT......G.T.TA....T..T.TG..T..GCAA...GGAA....GC..A.TG.CT.C

                    E
                910 |     920       930       940       950       960       970       980       990      1000
D2  TTCATCTTACTGACAGCTGTCACTCCTTCAATGACAATGCGTTGCATAGGAATGTCAAATAGAGACTTTGTGGAAGGGGTTTCAGGAGGAAGCTGGGTTG
    F  I  L  L  T  A  V  T  P  S  M  T  M  R  C  I  G  M  S  N  R  D  F  V  E  G  V  S  G  G  S  W  V  D

D2  TTCATCTTACTGACAGCTGTCACTCCTTCAATGACAATGCGTTGCATAGGAATGTCAAATAGAGACTTTGTGGAAGGGGTTTCAGGAGGAAGCTGGGTTG
D1  .....T..G....TGCTG..A..A..A..T...G.C.....A...G.G.....AGGC..C.........C.........AC.G.......C..CA.....G.
D3  ..T..AC..T.A.TGCTG..T..C..A..C.........A.A..TG.....G.AGG...C.....T...........CC.A..G....CT.CG.......
D4  ..TG..C..A...TGCTG...G.C..A..CTACGG......A...G.....G.AGG...C.................A..C.....T...GCA.....C.
```

FIGURE 31C

```
          1010       1020       1030       1040       1050       1060       1070       1080       1090       1100
D2  ACATAGTCTTAGAACATGGAAGCTGTGTGACGACGATGGCAAAAAACAAACCAACATTGGATTTTGAACTGATAAAAACAGAAGCCAAACAGCCTGCCAC
      I  V  L  E  H  G  S  C  V  T  T  H  A  K  N  K  P  T  L  D  F  E  L  I  K  T  E  A  K  Q  P  A  T

D   ACATAGTCTTAGAACATGGAAGCTGTGTGACGACGATGGCAAAAAACAAACCAACATTGGATTTTGAACTGATAAAAACAGAAGCCAAACAGCCTGCCAC
D   .TG.G..AC.G..G........T..C..C..C..C.....................C....CA.......CT.G..G..G..G.T..C.A.C.....AGT
D   ..G.G..GC.C..G..C..TG.G........T..C.....T..G.....G..C..GC....CA.A..G..TCAG..G..C..G....CC..A.TG..G..
D   .TC.G..GC...........G.A..C..C..A..C.....CC.GGGA........C.................CT..G...AC......GG.AGTG..TCT


          1110       1120       1130       1140       1150       1160       1170       1180       1190       1200
D2  CCTAAGGAAGTACTGTATAGAGGCAAAGCTAACCAACACAACAACAGAATCTCGCTGCCCAACACAAGGGGAACCCAGCCTAAATGAAGAGCAGGACAAA
      L  R  K  Y  C  I  E  A  K  L  T  N  T  T  T  E  S  R  C  P  T  Q  G  E  P  S  L  N  E  E  Q  D  K

D2  CCTAAGGAAGTACTGTATAGAGGCAAAGCTAACCAACACAACAACAGAATCTCGCTGCCCAACACAAGGGGAACCCAGCCTAAATGAAGAGCAGGACAAA
D1  T..GC.T..A.TG..C..T..A..T..AA..T.A.....C..C..C..T..GA.A..T...........A...G...CA..GGTG.....A..A...GCG
D3  ..........CTA..C..T....G...AA.T.......T......C..C..AA.A..T..C..C.........G.G.TTT..CC...G.........C.G
D4  GT....A.CC..T..C..T..A..CTCAA..T.A....T...C..G.C.A.AA.A..T.....G.....A..G..TTAT.....A..G..A..A...C..


          1210       1220       1230       1240       1250       1260       1270       1280       1290       1300
D2  AGGTTCGTCTGCAAACACTCCATGGTAGACAGAGGATGGGGAAATGGATGTGGACTATTTGGAAAGGGAGGCATTGTGACCTGTGCTATGTTCAGATGCA
      R  F  V  C  K  H  S  M  V  D  R  G  W  G  N  G  C  G  L  F  G  K  G  G  I  V  T  C  A  M  F  R  C  K

D2  AGGTTCGTCTGCAAACACTCCATGGTAGACAGAGGATGGGGAAATGGATGTGGACTATTTGGAAAGGGAGGCATTGTGACCTGTGCTATGTTCAGATGCA
D1  .AC..T..G...CG..GAA.GT.C..G.........C.....C.....C.....G.....C.....A..TA.TC.AA.A..G.....C.A...T.AG..TG
D3  .AC.A...G..T..G..TA.ATAC..G.........C.........C..T.....TT.G.....C......A..T.G.....A..C..G.AA..TCA...TT
D4  CA..A.A.T...CGGAGAGATG.............G.....C.....C.....CT.G.........A.....AG.......A.....G.A...TTC...TT


                                                                                SphI
          1310       1320       1330       1340       1350       1360       1370       1380       1390       1400
D2  AAAAGAACATGGAAGGAAAAGTTGTGCAACCAGAAAACTTGGAATACACCATTGTGATAACACCTCACTCAGGGGAAGAGCATGCAGTCGGAAATGACAC
      K  N  M  E  G  K  V  V  Q  P  E  N  L  E  Y  T  I  V  I  T  P  H  S  G  E  E  H  A  V  G  N  D  T

D2  AAAAGAACATGGAAGGAAAAGTTGTGCAACCAGAAAACTTGGAATACACCATTGTGATAACACCTCACTCAGGGGAAGAGCATGCAGTCGGAAATGACAC
D1  TG.CA..AC.A........GA.A.CT...TAT......C.AA....TT.AG.GA.AG.C..CGTC...A.T..A..TC....CCAG..G........G..
D3  T.G.ATCA..A..G........G.......AT..G...C.CA.........G.CA.C..C...GTG...A....A..CC.A..CCAG..G........A..
D4  CGGG...G..AAC...C..TT.G..C...ATT..G...C.T........AG.G..TG.....GTC...AAT..A..CACC........A...........
```

FIGURE 31D

```
          1410       1420       1430       1440       1450       1460       1470                  1490       1500
D2  AGGAAAACATGGCAAGGAAATCAAAATAACACCACAGAGTTCCATCACAGAAGCAGAATTGACAGGTTATGGCACTGTCACAATGGAGTGCTCTCCAAGA
     G  K  H  G  K  E  I  K  I  T  P  Q  S  S  I  T  E  A  E  L  T  G  Y  G  T  V  T  M  E  C  S  P  R

D2  AGGAAAACATGGCAAGGAAATCAAAATAACACCACAGAGTTCCATCACAGAAGCAGAATTGACAGGTTATGGCACTGTCACAATGGAGTGCTCTCCAAGA
D1  TAC.G.......A.CAACTGCA.CC........T..AGC.C.T.CGT.G...AT.C.GC....C.AC..C..A..CC.T...T.A..T..T..A..T..G
D3  GCAGGG.GTCACGGCT..G..A------.....C...GCA..A.C.G.T.....CATT..AC.T.AA.....A..CC..GGGC.A..A.....A...C.G
D4  .TCC..T.....AGTTAC.GC..CG.....T..CAG.TCAC.ATCGGTG....TCA.....C.G.AC.....AGAAC.A...C.C..T..TGAA..C..G

                    1520       1530       1540       1550       1560       1570       1580       1590       1600
D2  ACGGGCCTCGACTTCAATGAGATGGTGTTGCTGCAGATGGAAAATAAAGCTTGGCTGGTGCACAGGCAATGGTTCCTAGACCTGCCGTTACCATGGTTGC
     T  G  L  D  F  N  E  M  V  L  L  Q  M  E  N  K  A  W  L  V  H  R  Q  W  F  L  D  L  P  L  P  W  L  P

D2  ACGGGCCTCGACTTCAATGAGATGGTGTTGCTGCAGATGGAAAATAAAGCTTGGCTGGTGCACAGGCAATGGTTCCTAGACCTGCCGTTACCATGGTTGC
D1  ..A..G..A..T..T..C..............ACA...A....G...T.A.....T..C....AA..G.....T......T.A..AC.G..T...ACCT
D3  ..A..TT.G..T........A...A.C.CAT..ACA...A.G..C.....A...A....A..T..A.........T.T...T.A..CC........ACAT
D4  T.T..AA.T.....T.........A.TC..A..A.A...A....G...A.A.....T.....T.A.........TT.G..T..A..TC........ACAG

          1610       1620       1630       1640       1650       1660       1670       1680       1690       1700
D2  CCGGAGCGGACACACAAGGGTCAAATTGGATACAGAAAGAGACATTGGTCACTTTCAAAAATCCCCATGCGAAGAAACAGGATGTTGTTGTTTTAGGATC
     G  A  D  T  Q  G  S  N  W  I  Q  K  E  T  L  V  T  F  K  N  P  H  A  K  K  Q  D  V  V  V  L  G  S

D2  CCGGAGCGGACACACAAGGGTCAAATTGGATACAGAAAGAGACATTGGTCACTTTCAAAAATCCCCATGCGAAGAAACAGGATGTTGTTGTTTTAGGATC
D1  .T..G..TTTA...TCCCAAGAG.C.....ACAGAC....TTT.C.......A..T..G.CAG.T.....A.....G.....A..A..C..AC.......
D3  .A.....TACAG..G..ACAC...C.....ACAG.......CTTC.T..G..A..T......G.A.....A..A..G..A..A..A......C.T.....
D4  .A.....A......TC..A.GTTC.C....ATT.C.......G.A....G..A..T..GGT...T.....C....G.........GACA..GC.G.....

          1710       1720       1730       1740       1750       1760       1770       1780       1790       1800
D2  CCAAGAAGGGGCCATGCACACAGCACTTACAGGGGCCACAGAAATCCAAATGTCATCAGGAAACTTACTCTTCACAGGACATCTCAAGTGCAGGCTGAGA
     Q  E  G  A  M  H  T  A  L  T  G  A  T  E  I  Q  M  S  S  G  N  L  L  F  T  G  H  L  K  C  R  L  R

D2  CCAAGAAGGGGCCATGCACACAGCACTTACAGGGGCCACAGAAATCCAAATGTCATCAGGAAACTTACTCTTCACAGGACATCTCAAGTGCAGGCTGAGA
D1  A........A..A........T..G..G..T..A..G.............C....GG.ACG.CAAC.A.T...G.......C..A..A.....A..A.A.
D3  A.....G..A..A.....T........G.....A..T.....G.......CC...GG...C.CAAGTA....TG.G..G..CT.A..A..T..A..C.AG
D4  T..G.....A........TT.T..C..CG.T..A.........G.GG.CTCCGGTGAT.....TCACA.G..TG..................AAG.CC.T
```

FIGURE 31E

```
          1810       1820       1830       1840       1850       1860       1870       1880       1890       1900
D2  ATGGACAAGCTACAGCTCAAAGGAATGTCATACTCTATGTGCACAGGAAAGTTTAAAGTTGTGAAGGAAATAGCAGAAACACAACATGGAACAATAGTTA
    M  D  K  L  Q  L  K  G  M  S  Y  S  M  C  T  G  K  F  K  V  V  K  E  I  A  E  T  Q  H  G  T  I  V  I

D2  ATGGACAAGCTACAGCTCAAAGGAATGTCATACTCTATGTGCACAGGAAAGTTTAAAGTTGTGAAGGAAATAGCAGAAACACAACATGGAACAATAGTTA
D1  .........A...ACTT.A.....G.........TGTG...........CTCA..C..GT.A.A...A...G.G..T..G..C..G.........TG.TC.GG
D3  .........AT.GG.A........G...AGC..TG.A......TTG..T.GC...GTGT.GAA...A...G.CT.C.....G..G.....G.......C.C.
D4  .....G..AT.GAGAA....G.............A.G.....TT..........CTC.A...AC..A..G..G...........G.....G....C...GG

          1910       1920       1930       1940       1950       1960       1970       1980       1990       2000
D2  TCAGAGTGCAATATGAAGGGGACGGCTCTCCATGCAAGATCCCTTTTGAGATAATGGATTTGGAAAAAAGACATGTCTTAGGTCGCCTGATTACAGTCAA
      R  V  Q  Y  E  G  D  G  S  P  C  K  I  P  F  E  I  M  D  L  E  K  R  H  V  L  G  R  L  I  T  V  N

D2  TCAGAGTGCAATATGAAGGGGACGGCTCTCCATGCAAGATCCCTTTTGAGATAATGGATTTGGAAAAAAGACATGTCTTAGGTCGCCTGATTACAGTCAA
D1  .GCAG..TA..........AACA.A.G.A...........T..C...TC..CCCAA...GA.A..GG.GC.ACCCAGAAT..GA.AT.A..A....C...
D3  .T.AG..TG.G..CA.....A.A.ATG.A..C........T.....CTCC.CGGA....GGAC..GG..A.GC.CA.AAT..CA.A.....C....C...
D4  .G.A...CA.G........T.CT..AG....G..T..AG....CA.A.......GA...G..A.C..G.A.A.A..GG.T..G..TA.C..CT..TC..C

                                                                T
          2010       2020       2030       2040       2050       2060       2070       2080       2090       2100
D2  CCCAATTGTGACAGAAAAAGATAGCCCAGTCAACATAGAAGCAGAACCTCCATTCGGAGACAGCTACATCATCATAGGAGTAGAGCCGGGACAACTGAAG
    P  I  V  T  E  K  D  S  P  V  N  I  E  A  E  P  P  F  G  D  S  Y  I  I  I  G  V  E  P  G  Q  L  K

D2  CCCAATTGTGACAGAAAAAGATAGCCCAGTCAACATAGAAGCAGAACCTCCATTTGGAGACAGCTACATCATCATAGGAGTAGAGCCGGGACAACTGAAG
D1  ...C..A..C..T..C.....A.AA........T..T..G........A..C.....T..G.........G.GG......C..GTGAAAA.GCTT....A
D3  T...G.G.....CA.G..G..GGAG..T........T..G..T........T........A..TA....AG.A..T...A.T.GAGACAA.GCC.....A
D4  ...TT.G.CTGAGA.T.CCA.C..TG..AC.........GTT......C..C.....G...........AG.G.....T..T.GAAACA.TGC.T.A.CA

          2110       2120       2130       2140       2150       2160       2170       2180       2190       2200
D2  CTCAACTGGTTTAAGAAAGGAAGTTCTATCGGCCAAATGTTTGAGACAACAATGAGGGGGGCGAAGAGAATGGCCATTTTAGGTGACACAGCCTGGGATT
    L  N  W  F  K  K  G  S  S  I  G  Q  M  F  E  T  T  M  R  G  A  K  R  M  A  I  L  G  D  T  A  W  D  F

D2  CTCAACTGGTTTAAGAAAGGAAGTTCTATCGGCCAAATGTTTGAGACAACAATGAGGGGGGCGAAGAGAATGGCCATTTTAGGTGACACAGCCTGGGATT
D1  ..A.G......C...........CAGC..A..GA..........AG....TGCCC.A..A..ACGA..G.........C.G..A.....C..A.....C.
D3  A.........AC.....G.....C..G..T..GA.G.....C...G.T..TGCC..A..T..A.G.C.C........C..G..A..............C.
D4  ...C.T.....C.G......G.....C..T...A.G.........T.C...TAC..A..T..A..AC...........C.......A.....T.......
```

FIGURE 31F

```
          2210       2220       2230       2240       2250       2260       2270       2280       2290       2300
D2  TTGGATCCTTGGGAGGAGTGTTTACATCTATAGGAAAGGCTCTCCACCAAGTCTTTGGAGCAATCTATGGAGCTGCCTTCAGTGGGGTTTCATGGACTAT
      G  S  L  G  G  V  F  T  S  I  G  K  A  L  H  Q  V  F  G  A  I  Y  G  A  A  F  S  G  V  S  W  T  M

D2  TTGGATCCTTGGGAGGAGTGTTTACATCTATAGGAAAGGCTCTCCACCAAGTCTTTGGAGCAATCTATGGAGCTGCCTTCAGTGGGGTTTCATGGACTAT
D1  .C..T..TA.A..........C..G.....G.....ACTGG.A.....G..T......A.TGCA.......T.TTG..T..C..A.....T.....C..
D3  .......AG....T..T..T..G.AT..AT....G..AATGG........A.A.....GAGTGCT..CAC...CCTA..TG....A..C..C....TG..
D4  ....T...G.T..T...C....C.....AT.G.........G.G.....G..T......AGTG.G...AC.A.CATG..TG.A..A..C.......TG..

          2310       2320       2330       2340       2350       2360       2370       2380       2390       2400
D2  GAAAATCCTCATAGGAGTCATTATCACATGGATAGGAATGAATTCACGCAGCACCTCACTGTCTGTGACACTAGTATTGGTGGGAATTGTGACACTGTAT
      K  I  L  I  G  V  I  I  T  W  I  G  M  N  S  R  S  T  S  L  S  V  T  L  V  L  V  G  I  V  T  L  Y

D2  GAAAATCCTCATAGGAGTCATTATCACATGGATAGGAATGAATTCACGCAGCACCTCACTGTCTGTGACACTAGTATTGGTGGGAATTGTGACACTGTAT
D1  ......AGGA.....GA.TC.GC.G......C.....T.A......A.G.A...G..C..T..G....TGTGCA.CGCA..T..C..G..C.........C
D3  ......TGGA.....T...C.CT.A..C........GT....C...AAA.AT..T..TA....AT.TT..TGCA.CGC.A.A.....CA.T.....C...
D4  T.G......A..T..GT..C.AG.GTTG.....T..C.C...C...A.G.A...T...A..G..A....GTGCA..GCT..T...GGAA.C..T....T.

                                  NS1
          2410       2420       2430       2440       2450       2460       2470       2480       2490       2500
D2  TTGGGAGTCATGGTGCAGGCCGATAGTGGTTGCGTTGTGAGCTGGAAAAACAAAGAACTGAAATGTGGCAGTGGGATTTTCATCACAGACAACGTGCACA
      L  G  V  M  V  Q  A  D  S  G  C  V  V  S  W  K  N  K  E  L  K  C  G  S  G  I  F  I  T  D  N  V  H  T

D2  TTGGGAGTCATGGTGCAGGCCGATAGTGGTTGCGTTGTGAGCTGGAAAAACAAAGAACTGAAATGTGGCAGTGGGATTTTCATCACAGACAACGTGCACA
D1  C.A..........T.....A...TCG..A..T..AA.C.A.......GG..G......T.........A..C..C.....TG....TA.TG.A..T....
D3  C......C.G.......A..T..C.TG..G..T..CA.A.A.......GG.........C.........A.....A.......
D4  C....CT...CA..T..A..A..G.TG.....T..G...TCA....GTGGG......T...GG
```

FIGURE 31G

```
                                                                                A
          2510       2520       2530       2540       2550       2560       2570                  2590       2600
D2  CATGGACAGAACAATACAAGTTCCAACCAGAATCCCCTTCAAAACTAGCTTCAGCTATCCAGAAAGCCCATGAAGAGGGCATTTGTGGAATCCGCTCAGT
      W  T  E  Q  Y  K  F  Q  P  E  S  P  S  K  L  A  S  A  I  Q  K  A  H  E  E  G  I  C  G  I  R  S  V
                                                                                D

D2  CATGGACAGAACAATACAAGTTCCAACCAGAATCCCCTTCAAAACTAGCTTCAGCTATCCAGAAAGCCCATGAAGAGGACATTTGTGGAATCCGCTCAGT
D1  .T........G........A.....GG.T..C.....CAAG.G....T.AG....C..TGG...G..ATGG..G....GTG.G
D3
D4
```

```
                                                                                HpaI
          2610       2620       2630       2640       2650       2660       2670       2680       2690       2700
AACAAGACTGGAGAATCTGATGTGGAAACAAATAACACCAGAATTGAATCACATTCTATCAGAAAATGAGGTGAAGTTAACTATTATGACAGGAGACATC
  T  R  L  E  N  L  M  W  K  Q  I  T  P  E  L  N  H  I  L  S  E  N  E  V  K  L  T  I  M  T  G  D  I
```


FIGURE 31H

```
       Capsid
             10        20        30        40        50        60        70        80        90       100
D2  MNNQRKKAKNTPFNMLKRERNRVSTVQQLTKRFSLGMLQGRGPLKLFMALVAFLRFLTIPPTAGILKRWGTIKKSKAINVLRGFRKEIGRMLNILNRRRR
D1                                    ...K.L.S.Q..M..V..FI.....A........A...SF..NG..K.....KR..SN....M...K.
D3                                                                                K..K..K...SN..S.I.K.KK
D4                                                                 ......QL..N...RI.I.............G.K.

                  prM                                                                    *
           110       120       130       140       150       160       170       180     |  190       200
D2  SAGMIIMLIPTVMAFHLTTRNGEPHMIVSRQEKGKSLLFKTEVGVNMCTLMAMDLGELCEDTITYKCPLLRQNEPEDIDCWCNSTSTWVTYGTCTTMGEH
D1  .VT.LL..L..AL.......G........K..R........SA......I...........M....RITEA..D.V.....A.D........SQT...
D3  TSLCLM.ML.ATL.....S.D...R...GKN.R........AS.I.....I......M.D..V....HITEV..........L..........NQA...
D4  .TITLLC..........S..D...L...AKH.R.RP.....TE.I.K...I......M...V.......VNT..........L.........Y.QS..R

        M                                                                      E
                   220       230       240       250       260       270               290       300
D2  RREKRSVALVPHVGMGLETRTETWMSSEGAWKHVQRIETWILRHPGFTMMAAILAYTIGTTHFQRALIFILLTAVTPSMTMRCIGMSNRDFVEGVSGGSW
D1  ..D......A....L.....A...........QI.KV...A.......VI.LF..HA...SIT.KGI.....ML.....A...V.IG.......L..AT.
D3  ..D......A.......D...Q....A....RQ.EKV...A.......IL.LF..HY...SLT.KVV.....ML.........V.VG.......L..AT.
D4  .........T..S.......A.............A..V.S....N...ALL.GFM..M..Q.GI..TVF.V.MML.A..YG...V.VG...........A.

                                                *
           310       320       330       340    |            360       370       380       390       400
D2  VDIVLEHGSCVTTMAKNKPTLDFELIKTEAKQPATLRKYCIEAKLTNTTTESRCPTQGEPSLNEEQDKRFVCKHSMVDRGWGNGCGLFGKGGIVTCAMFR
D1  ..V..................I..L...VTN..V...L.....IS....D.........AT.V....AN...RRTF...............SLI...K.K
D3  ..V.....G............I..Q....T.L.....L...G.I..I..D.........AI.P....QNY....TY................SLI...K.K
D4  ..L.....G......QG.........T..T..EV.L..T.....SIS.I..AT.........Y.K....QQYI.RRDV................V....K.S

                                   *
           410       420       430 |      440       450       460       470       480       490       500
D2  CKKNMEGKVVQPENLEYTIVITPHSGEEHAVGNDTGKHGKEIKITPQSSITEAELTGYGTVTMECSPRTGLDFNEMVLLQMENKAWLVHRQWFLDLPLPW
D1  .VTKL...IA.Y...K.SVIV.V.T.DQ.Q...E.TE..TTAT....APTS.IQ..D...L.LD...............T.KK.S....K..........
D3  .LESI.......H...K..VI..V.T.DQ.Q...E.QGVTA..--...A.TA..I.PE...LGL..............IS.T.K....M......F.....
D4  .SGKIT.NL..I......V.V.V.N.DT.......SN..VTAT...R.PSV.VK.PD..EL.LD.E..S.I.....I.MK.KK.T....K..........
```

FIGURE 32A

```
          510       520       530       540       550       560       570       580       590       600
D2  LPGADTQGSNWIQKETLVTFKNPHAKKQDVVVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKCRLRMDKLQLKGMSYSMCTGKFKVVKEIAETQHGTI
D1  TS..L.SQET.NRQDL.....TA.....E.......................................................................
D3  TS..TAETPT.NR..L......A.....E.....................T.GTTTI.A.......K....T......V....S..LE..V.......V
D4  TA....SEVH.NY..RM....V....R...T..........S..A....VDSGD..HM.A.....KV..E..RI.....T..S...SID..M.......T

          610       620       630       640       650       660       670       680       690       700
D2  VIRVQYEGDGSPCKIPFEIMDLEKRHVLGRLITVNPIVTEKDSPVNIEAEPPFGDSYIIIGVEPGQLKLNWFKKGSSIGQMFETTMRGAKRMAILGDTAW
D1  LVQ.K...TDA......STQ.EKGATQN.....A.....D.EK.....................................................
D3  L.K.E.K.KDA......STE.GQGKAHN.....A..V..K.EE...........E...VV.AGEKA...S.........K...A.A...R..........
D4  .VK.K...A.A...V.I..R.VN.KK.V..I.SST.LAENTN.AT...L.........V...GNSA.T.H..R......K...S.Y..........E...

          710       720       730       740       750       760       770       NS1       790       800
D2  DFGSLGGVFTSIGKALHQVFGAIYGAAFSGVSWTMKILIGVIITWIGMNSRSTSLSVTLVLVGIVTLYLGVMVQADSGCVVSWKNKELKCGSGIFITDNV
D1  ....I......M..LV.....TA..VL..........G..ILL..L.L...N.....MCIA..M.....................................
D3  ....V...LN.L..MV..I..SA.T.L.G....M...G...LL....L..KN..M.FSCIAI..I.....AV....M...IN..GR.........V.NE.
D4  ....V..L...L...V.....SV.TTM.G....MIR....FLVL...T...N..MAM.CIA..GI..F..FT...EM......SG...R

          810       820       830       840       850       860       870       880       890       900
D2  HTWTEQYKFQPESPSKLASAIQKAHEEDICGIRSVTRLENLMWKQITPELNHILSENEVKLTIMTGDIKGIMQAGKRSLRPQPTELKYSWKTWGKAKMLS
D1  ..........AD..KR.SA..G..W..GV
D3
D4
```

FIGURE 32B

```
                                                    prM
          410        420        430                 450        460        470        480        490
CAGGCATGATCATTATGCTGATTCCAACAGTGATGGCGTTCCATTTAACCACACGTAACGGAGAACCACACATGATCGTCAGCAGACAAG
 G  M  I  I  M  L  I  P  T  V  M  A  F  H  L  T  T  R  N  G  E  P  H  M  I  V  S  R  Q  E

                                                F  H  L  S  S  R
                                                TTCCATCTCTCGAGCCGT
                                                    CTCGAG
                                                    XhoI
                                                F  H  L  S  S  R
D2.VAC      CAGGCATGATCATTATGCTGATTCCAACAGTGATGGCGTTCCATCTCTCGAGCCGTAACGGAGAACCACACATGATCGTCAGCAGACAAG
D1.VAC      TGAC....C..C........C.G..C....CCC.....TTCCATCTCTCGAGCCGAGGG.....G..G..T.....A..T....AG..G.
D3.VAC      .GCT.TGTC....G...A..T.A...G..ACAC.T..T.....CC.CT.GAGCCGAG.T.....G..G.G......T..GG.G.AGA.T.
D4.VAC      .GATA.CAT.GC.GTGCT.......C..C..A........T..CC.CT.G.GCAGAG.T..C.....C.T......A..GGCA.A...T.
```

Amplicon DEN-2 F1-prM5':

Upstream Primer= D2-SMT71

Downstream Primer=cD2-prM/Xho/447 = 3'-CACTACCGCAAGGTAGAGAGCTCGGCATTGCCTCTTGGTG-5' (SEQ ID NO: 127)

Amplicon DEN-2 F1-prM3':

Upstream Primer= D2-prM/Xho/447 = 5'-GTGATGGCGTTCCATCTCTCGAGCCGTAACGGAGAACCAC-3' (SEQ ID NO: 128)

Downstream Primer=cD2-1503

Heterologous Amplicons for Chimeric Constructs:

Upstream Primer= D1-prM/Xho/447 = 5'-GCCCTGGCGTTCCATCTCTCGAGCCGAGGGGGAGAGCCGC-3' (SEQ ID NO: 129)

D3-prM/Xho/447 = 5'-ACACTTGCTTTCCACCTCTCGAGCCGAGATGGAGAGCCGC-3' (SEQ ID NO: 130)

D4-prM/Xho/447 = 5'-GTAATGGCGTTTCACCTCTCGAGCAGAGATGGCGAACCCC (SEQ ID NO: 131)

FIGURE 33

```
                2310       2320       2330       2340       2350       2360       2370       2380       2390
GAAAATCCTCATAGGAGTCATTATCACATGGATAGGAATGAATTCACGCAGCACCTCACTGTCTGTGACACTAGTATTGGTGGGAATTGT
 K  I  L  I  G  V  I  I  T  W  I  G  M  N  S  R  S  T  S  L  S  V  T  L  V  L  V  G  I  V
```

```
                    N  S  R  S
                    AATTCTAGAAGC
                    TCTAGA
                    XbaI
```

```
                                                N  S  R  S
D2.VAC      GAAAATCCTCATAGGAGTCATTATCACATGGATAGGAATGAATTCTAGAAGCACCTCACTGTCTGTGACACTAGTATTGGTGGGAATTGT
D1.VAC      .......AGGA.....GA.TC.GC.G......C.....T.A.....TA.A.A...G..C..T..G....TGTGCA.CGCA..T..C..G..
D3.VAC      ......TGGA.....T...C.CT.A..C........GT....C..TAGA.AT..T..TA....AT.TT..TGCA.CGC.A.A.....CA.
D4.VAC      T.G......A..T..GT..C.AG.GTTG.....T..C.C...C..TA.G.A...T...A..G..A....GTGCA..GCT..T...GGAA.
```

Amplicon DEN-2 F2-E5':

Upstream Primer= D2-1261

Downstream Primer=cD2-E/Xba/2344 = 3'-CCTATCCTTACTTAAGATCTTCGTGGAGTGACAGAC-5' (SEQ ID NO: 132)

Amplicon DEN-2 F2-E3':

Upstream Primer= D2-E/Xba/2344 = 5'-GGATAGGAATGAATTCTAGAAGCACCTCACTGTCTG-3' (SEQ ID NO: 133)

Downstream Primer=cD2-2704 = 5'-CCGCAGAGATCGTTTTCCTGCCTGCATGATTCC-3' (SEQ ID NO: 134)

Heterologous Amplicons for Chimeric Constructs:

Downstream Primer= D1-E/Xba/2344 = 3'-CCGATCCTAATTTAAGATCTTTGTGCAGGGAAAGCC-5' (SEQ ID NO: 135)

D3-E/Xba/2344 = 3'-CCTATCCCAACTTGAGATCTTTATGAAGATACAGTA-5' (SEQ ID NO: 136)

D4-E/Xba/2344 = 3'-CCTAACCGTGCTTGAGATCTTTGTGAAGTTACCGAC-5' (SEQ ID NO: 137)

FIGURE 34A

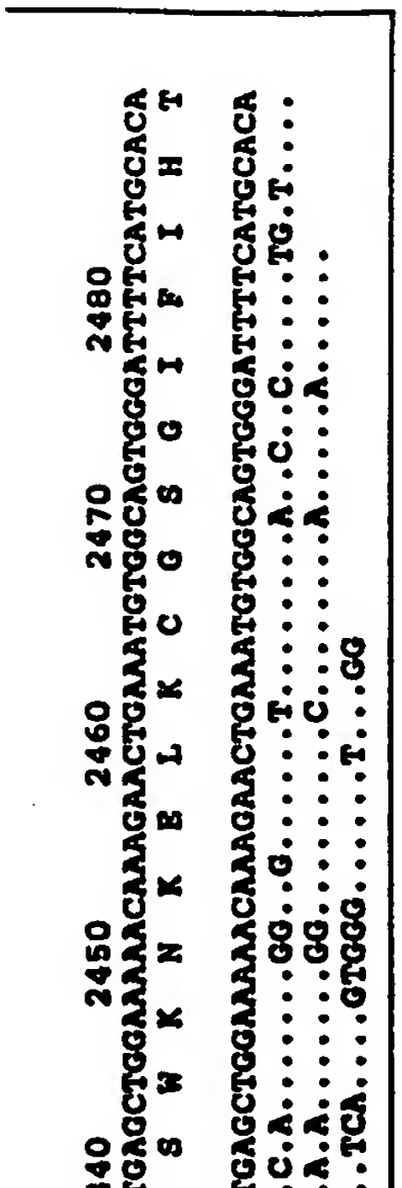

NS1
2410 2420 2430 2440 2450 2460 2470 2480
TTGGGAGTCATGGTGCAGGCCGATAGTGGTTGCGTTGTGAGCTGGAAAACAAGAACTGAAATGTGCACA
L G V M V Q A D S G C V V S W K N K E L K C G S G I F I H T
D2.VAC TTGGGAGTCATGGTGCAGGCCGATAGTGGTTGCGTTGTGAGCTGGAAAACAAGAACTGAAATGTGCACA
D1.VAC C.A.....T.TCG..A..T..AA.C.A.....GG..G.....T.....A...TG.T....
D3.VAC C.....C.G.....T..G..T..CA.A.A.....GG.....C.....A..T..A..A..G.TG.....T..G...TCA....GTGGG.....T...GG
D4.VAC C....CT...CA..T

FIGURE 34B

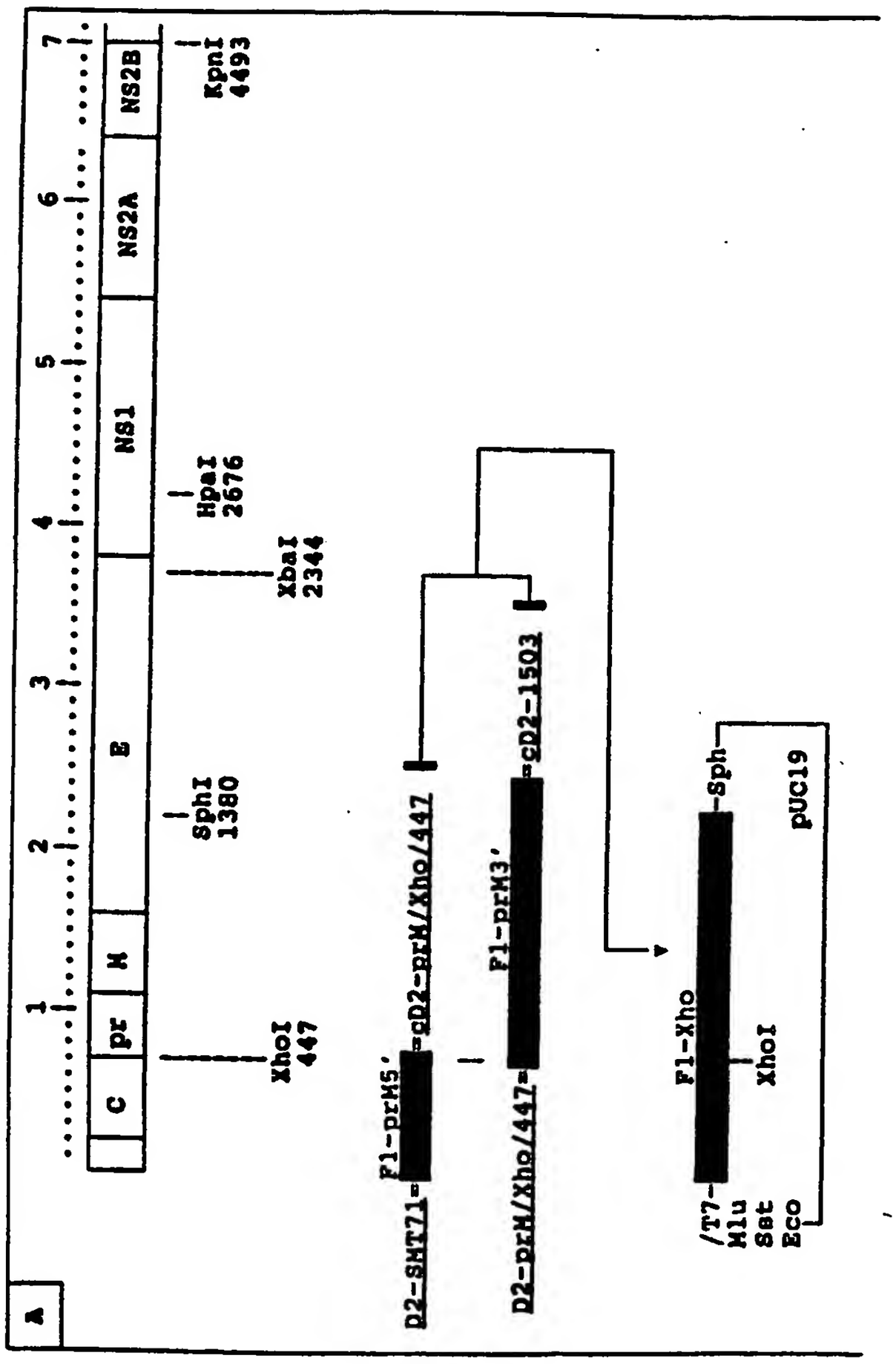

A
C
pr
M
E
NS1
NS2A
1
2
3
4
5
6
7
XhoI 447
SphI 1380
XbaI 2344
HpaI 2676
KpnI 4493
D2-SHTZ1=F1-prM5'
D2-prM/Xho/447=F1-prM3'
=cD2-1503
F1-Xho
/T7-
Mlu
Sst
Eco
XhoI
Sph
pUC19

FIGURE 35A

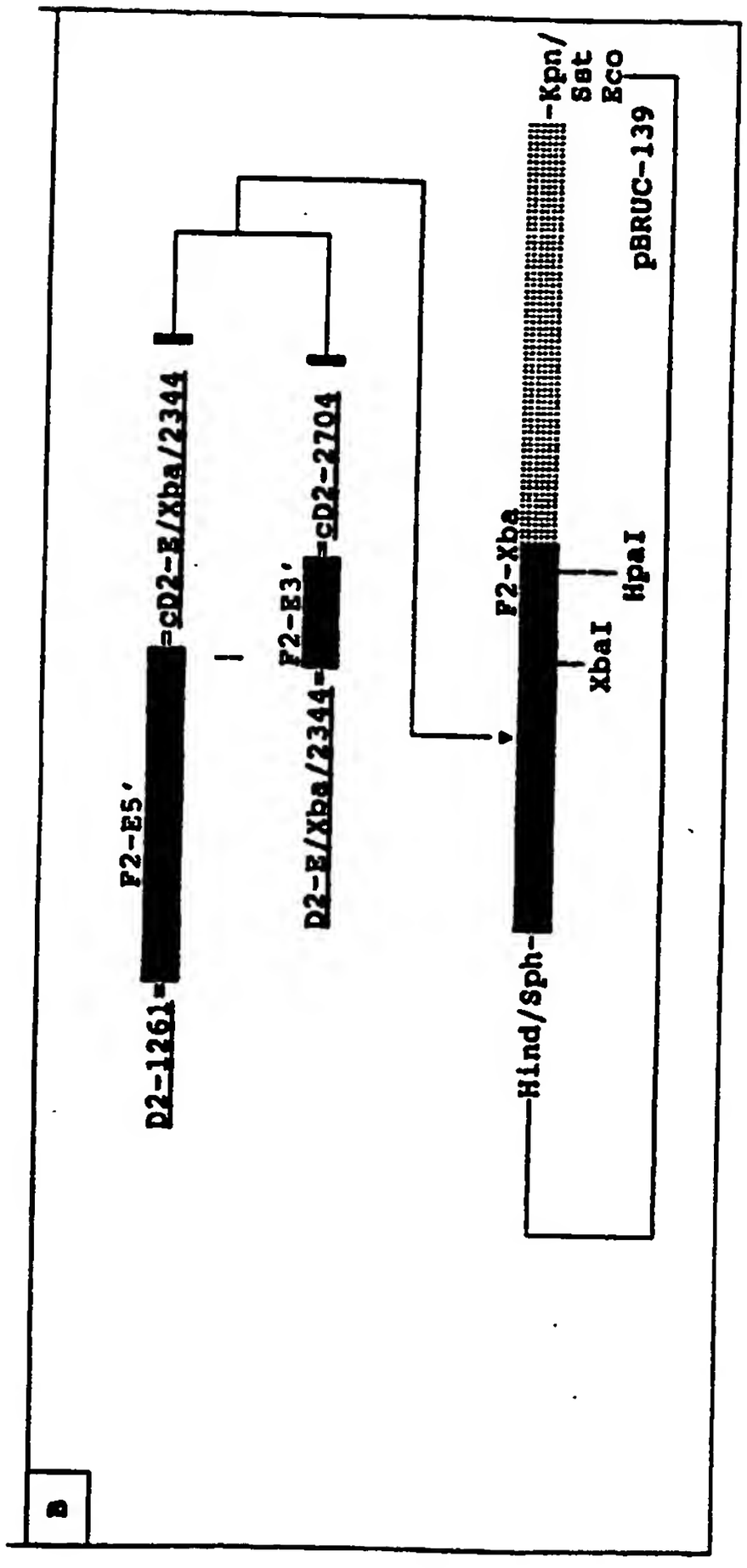

D2-1261
F2-E5'
=cD2-E/Xba/2344
D2-E/Xba/2344
F2-E3'
=cD2-2704
Hind/Sph-
F2-Xba
XbaI
HpaI
-Kpn/
Sst
Eco
pBRUC-139

FIGURE 35B

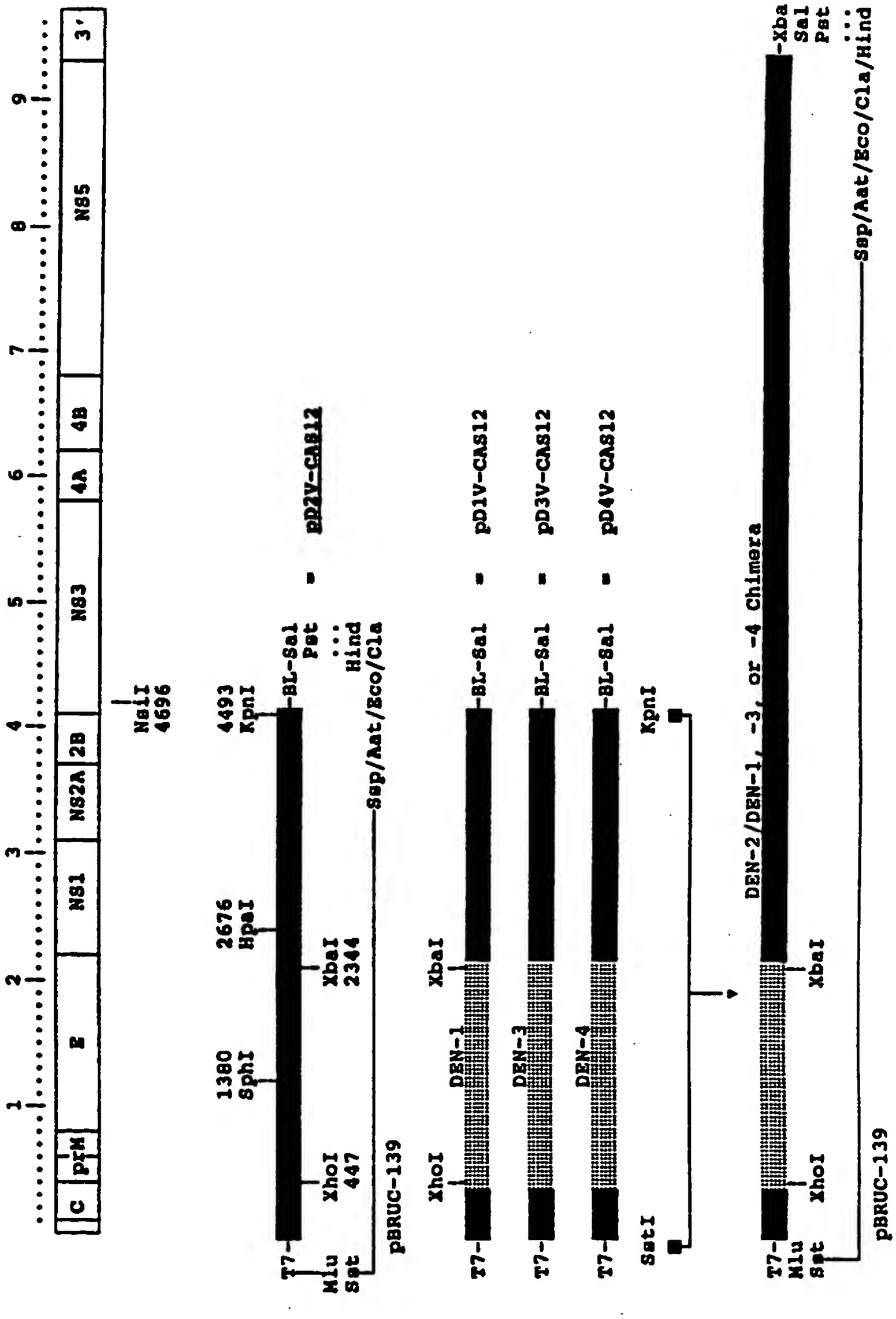

1
2
3
4
5
6
7
8
9
C
PrM
E
NS1
NS2A
2B
NS3
4A
4B
NS5
3'
NaiI
4696
4493
KpnI
2676
HpaI
XbaI
2344
1380
SphI
XhoI
447
T7-
Mlu
Spt
pBRUC-139
BL-Sal
Pst
...
Hind
Sap/Aat/Eco/Cla
= pD2V-CAS12
XbaI
DEN-1
XhoI
= pD1V-CAS12
DEN-3
= pD3V-CAS12
DEN-4
= pD4V-CAS12
KpnI
SstI
DEN-2/DEN-1, -3, or -4 Chimera
-Xba
Sal
Pst
...
-Sap/Aat/Eco/Cla/Hind

FIGURE 36

INTERNATIONAL SEARCH REPORT

Intr tional Application No
PCT/US 96/09209

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/40 C12N15/86 C07K14/18 A61K39/12 C12N7/01
C12N7/00 C12N5/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	VACCINE, vol. 14, no. 4, March 1996, GUILDFORD GB, pages 329-336, XP000579824 VAUGHN, D.W. ET AL.: "Testing of a Dengue 2 live-attenuated vaccine (strain 16681 PDK 53) in ten american volunteers" see the whole document ---	1
X	VIROLOGY, vol. 187, no. 4, April 1992, ORLANDO US, pages 573-590, XP000601641 BLOK, J. ET AL.: "Comparaison of Dengue -2 virus and its candidate vaccine derivative: sequence relationships with the Flaviviruses and other viruses"	7-12, 22-29,36
Y	see the whole document ---	1
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
6 September 1996	23. 10. 96
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	**Authorized officer** Chambonnet, F

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int ional Application No
PCT/US 96/09209

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO,A,93 06214 (US ARMY) 1 April 1993 see claims 40,63-68	1
A	AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE, vol. 47, no. 4 sup, 1992, pages 99-100, XP000600344 VAUGHN, D.W. ET AL.: "Phase I testing of a dengue-2 live-attenuated vaccine strain 16681 PDK 53 in american volunteers" see the whole document & 41st Annual Meeting of the American Society of Tropical Medicine and Hygiene Washington, USA November 15-19 1992	1
A	WO,A,92 03161 (US GOVERNMENT) 5 March 1992 see the whole document	1
A	WO,A,93 22440 (UNIV SINGAPORE ;TAN YIN HWEE (SG); FU JIANLIN (SG); TAN BOON HUAN) 11 November 1993 see the whole document	1,2,5,6, 13
A	WO,A,92 03545 (VIROGENETICS CORP) 5 March 1992 see claims 1,9,10,16-23,26; example 13	1
A	VIROLOGY, vol. 174, no. 2, February 1990, ORLANDO US, pages 479-493, XP002012813 RICO-HESSE, R.: "Molecular evolution and distribution of Dengue Viruses type 1 and 2 in nature" see the whole document	1
A	JOURNAL OF GENERAL VIROLOGY, vol. 69, no. 6, June 1988, pages 1391-1398, XP000600928 GRUENBERG, A. ET AL.: "Partial nucleotide sequence and deduced amino acid sequence of the structural proteins of Dengue virus type 2, New Guinea C and PUO-218 strains" see the whole document	1

-/--

2

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 96/09209

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	VIROLOGY, vol. 162, no. 1, January 1988, ORLANDO US, pages 167-180, XP000600931 HAHN, Y.S. ET AL.: "Nucleotide sequence of Dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses" see the whole document -----	1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/09209

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 6
because they relate to subject matter not required to be searched by this Authority, namely:

Remark: Although this claim is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. [ ] As all searchable claims could be searches without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int ional Application No
PCT/US 96/09209

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9306214	01-04-93	AU-B-	667836	18-04-96
		AU-B-	2691492	27-04-93
		CA-A-	2116980	01-04-93
		EP-A-	0604566	06-07-94
		JP-T-	6511150	15-12-94
WO-A-9203161	05-03-92	AU-B-	8762591	17-03-92
		US-A-	5494671	27-02-96
WO-A-9322440	11-11-93	AU-B-	4257593	29-11-93
		CA-A-	2134666	11-11-93
		EP-A-	0638122	15-02-95
WO-A-9203545	05-03-92	US-A-	5514375	07-05-96
		AU-B-	657711	23-03-95
		AU-B-	8728791	17-03-92
		GB-A,B	2269820	23-02-94
		JP-T-	6503227	14-04-94